

# The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics

Thomas E. Ouldridge

*Department of Bioengineering, Imperial College London,  
South Kensington Campus, London, SW7 2AZ, UK.*

Improved understanding of molecular systems has only emphasised the sophistication of networks within the cell. Simultaneously, the advance of nucleic acid nanotechnology, a platform within which reactions can be exquisitely controlled, has made the development of artificial architectures and devices possible. Vital to this progress has been a solid foundation in the thermodynamics of molecular systems. In this pedagogical review and perspective, I will discuss how thermodynamics determines both the overall potential of molecular networks, and the minute details of design. I will then argue that, in turn, the need to understand molecular systems is helping to drive the development of theories of thermodynamics at the microscopic scale.

## I. INTRODUCTION

Thermodynamics was originally developed in the 19th Century, driven by the dawn of the industrial revolution [1], and a desire to understand and optimise the extraction of useful work from engines. This work could be harnessed to pump water out of mines, or drive locomotives, for example. Although these machines were mechanical devices powered by the flow of heat, the fundamental source of the work was the chemical fuel - typically coal, initially. Chemical processes were incorporated into the framework of thermodynamics by Gibbs and others [2], leading to an understanding of the spontaneity of, and energy exchanged during, chemical reactions and phase changes.

Thermodynamics as originally introduced was a theory based entirely on the interrelation of macroscopic observables, such as temperature, pressure, volume and energy. In the late 19th and early 20th centuries, the development of statistical mechanics provided a microscopic basis for the theory, explaining how these bulk properties emerge from microscopic system properties [3]. In the process, the concept of entropy – the mysterious quantity whose increase is responsible for the thermodynamic arrow of time – was made far more concrete, as a statistical measure of the uncertainty of the microscopic state of a system. The exploration of statistical mechanics led to the development of theories of critical phenomena; explaining the exotic yet often universal behaviour of systems as they approach certain kinds of phase transitions [4, 5]; statistical mechanics is also the fundamental tool underlying the field of molecular simulation [6, 7].

Statistical mechanics and thermodynamics are often introduced as the study of equilibrium states, in which there is no net tendency for the system to evolve over time unless driven from the outside. Nonetheless, from the earliest days, Boltzmann and others proposed theoretical descriptions for the evolution of non-equilibrium systems and their subsequent relaxation to equilibrium [3]. The developing field of stochastic thermodynamics, in which the probabilistic description underlying statistical mechanics is extended to describe trajectories of non-

equilibrium systems through state space, has recently provided remarkable understanding of systems arbitrarily far from equilibrium [8–10].

Over the same period, our understanding of biomolecular systems has been transformed from complete ignorance to the ability to rationally design synthetic circuits and self-assembling architectures *in vitro* and *in vivo*. Indeed, although proteins were identified as far back as the early 19th Century [11, 12], their role as enzymes in living organisms was not demonstrated until 1926 [13], and protein structures were first solved in 1958 [14, 15]. Similarly, the genetic information-carrying role of biological DNA was first demonstrated in 1944 [16], the double-helical structure solved in 1953 [17] and the central dogma of molecular biology first stated in 1958 [18, 19]. Since then, through advances in crystallography, microscopy and other technologies, the molecular mechanisms of an enormous biochemical processes have been identified. Additionally, systems biology has shown how individual component reactions can combine to provide the complex behaviour exhibited by cells [20] – although we remain far from a full understanding of such sophisticated systems.

In the process of understanding some of the molecular complexity of the cell, we have shown that it contains microscopic analogues of the mechanical engines of the 19th Century. Molecular motors such as myosin consume chemical fuel to generate locomotive forces [21], and enzymatic pumps consume the same fuel to drive ions across membranes [22]. These membranes then act as capacitors that provide an alternative supply of power, like batteries for electric motors.

A deep appreciation of natural biomolecular systems is worthwhile in and of itself, and it provides an important contribution to the advancement of medicine. But equally, this hard-won understanding has laid the groundwork for the engineering of artificial systems and devices. In synthetic biology, novel molecular circuitry is often built by connecting naturally-occurring or slightly mutated proteins via artificial transcriptional regulation pathways [23]. At the same time, the molecular nanotechnology community has constructed systems based

on artificial components, including non-biological DNA and RNA sequences [24] and even artificial proteins [25, 26]. At the interface of these communities are those who combine the functionality of synthetic and natural components, both *in vivo* and *in vitro* [27, 28]. In aggregate, this work has produced remarkable results, including nanoscale self-assembly [29–31], implementation of molecular computation and control architectures [32–35] and repurposing of microbes for industry and healthcare [23, 36, 37].

In this pedagogical perspective, I will first discuss the basics of traditional chemical thermodynamics as it applies to biomolecular systems. I will then outline some of the ways that these ideas shape our understanding and design of molecular systems, both at a fundamental and a practical level. Subsequently, I will briefly discuss the emerging field of stochastic thermodynamics. This extension of traditional thermodynamics to fluctuating, far-from equilibrium contexts finds its most natural application in the analysis of molecular systems. Indeed, I will then argue that the very process of exploring abstract thermodynamic ideas in concrete biomolecular systems is in turn providing a deeper understanding of the fundamental thermodynamics.

## II. FUNDAMENTALS OF STATISTICAL MECHANICS

### A. The partition function and thermodynamic quantities

A closed system has no influx of energy or matter, and will eventually relax to thermodynamic equilibrium – a condition in which there are no net flows of energy or matter within the system itself. The second law of thermodynamics states that a quantity known as the entropy will systematically increase during this relaxation, reaching its maximum value in the equilibrium limit. The basic assumption of classical statistical mechanics, the microscopic theory underlying equilibrium thermodynamics, is that equilibrium for a closed system corresponds to equal occupancy of all accessible (equal-energy) *microstates* [38]. Here, a microstate is specified by the positions and momenta  $\mathbf{x}, \mathbf{p}$  of all degrees of freedom of the system.

The principle of equal *a priori* probability implies that the microstate of a closed system is as uncertain (hard to predict) as possible in equilibrium, and that closed systems must tend towards this maximally uncertain distribution. These ideas suggest a microscopic interpretation of the second law of thermodynamics. If entropy is a statistical measure of uncertainty about the microstate occupied by a system, then the tendency of this uncertainty to increase over time as a closed system tends towards equilibrium is the underlying cause of the second law. This idea will be explored further in Section VIII.

In the molecular context we are usually interested in a

relatively small system  $\sigma$  thermally connected to a much larger environment  $\Sigma$ , rather than a system in total isolation. This larger environment might be the lab as a whole, or perhaps a water bath for elevated temperatures. In any case, we are typically not concerned with the details of this environment  $\Sigma$ , other than in its role as a source and sink of energy in the form of heat. Generally, we assume that the coupling is weak so that it is reasonable to separately consider the energies of  $\Sigma$  and  $\sigma$  [6, 38].

The combined system of  $\sigma + \Sigma$  remains a closed system, which can reach equilibrium. However, due to energy exchange between  $\sigma$  and  $\Sigma$ , the energy of each component fluctuates and microstates of  $\sigma$  with different energies  $E_\sigma(\mathbf{x}, \mathbf{p})$  can be accessed. From applying the principle of equal *a priori* probability to the combined system of  $\sigma + \Sigma$ , it is possible to show that energy is shared such that microstates of  $\sigma$  are occupied with a probability density [6, 7, 38]

$$P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}) = \rho \frac{\exp(-E_\sigma(\mathbf{x}, \mathbf{p})/k_B T)}{Z_\sigma}, \quad (1)$$

where the environmental heat bath sets the temperature  $T$ , and the partition function  $Z_\sigma$  normalizes the distribution:

$$Z_\sigma = \rho \int d\mathbf{x} d\mathbf{p} \exp(-E_\sigma(\mathbf{x}, \mathbf{p})/k_B T), \quad (2)$$

with  $\rho$  a constant density of microstates that cancels out during calculations, but ensures the correct dimensionality. Eq. 2 is the famous Boltzmann distribution [6, 7, 38]; its form arises from sharing energy between  $\sigma$  and  $\Sigma$  in such a way as to maximise overall uncertainty in the microstate. Atomistic models of molecular systems, such as AMBER and CHARMM [39, 40], are essentially semi-empirical energy functions  $E_\sigma(\mathbf{x}, \mathbf{p})$ . Coarse-grained models, such as Martini and oxDNA [41, 42], are attempts to capture the behaviour of a reduced set of key degrees of freedom with a similar energy model. In either case, the models are typically too complex to be analysed directly. Instead, simulation is used to sample microstates of  $\sigma$ , allowing the properties of the system to be explored.

From the partition function follow the key thermodynamic quantities: the (average) internal energy of  $\sigma$  [6, 7, 38]

$$U_\sigma^{\text{eq}} = \int_{\mathbf{x}, \mathbf{p}} d\mathbf{x}, d\mathbf{p} E_\sigma(\mathbf{x}, \mathbf{p}) P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}) = k_B T^2 \frac{\partial}{\partial T} \ln Z_\sigma; \quad (3)$$

the entropy of  $\sigma$  [43]

$$\begin{aligned} S_\sigma^{\text{eq}} &= -k_B \int_{\mathbf{x}, \mathbf{p}} d\mathbf{x}, d\mathbf{p} P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}) \ln(P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p})/\rho) \\ &= k_B T \frac{\partial}{\partial T} \ln Z_\sigma + k_B \ln Z_\sigma; \end{aligned} \quad (4)$$

and the free energy of  $\sigma$  [6, 7, 38]

$$F_\sigma^{\text{eq}} = U_\sigma^{\text{eq}} - T S_\sigma^{\text{eq}} = -k_B T \ln Z_\sigma. \quad (5)$$

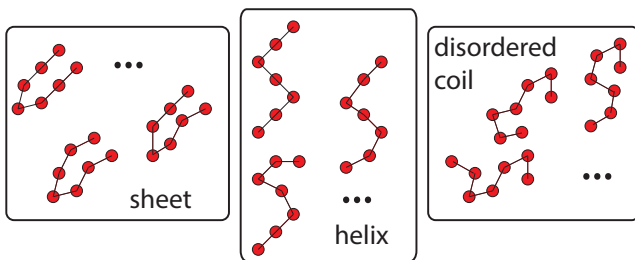


FIG. 1. Schematic illustration of the division of a continuous set of microstates into a relatively small number of macrostates. In this case there are three discrete macrostates, distinguished by the conformation of a small protein-like molecule. Similar conformations are grouped together and characteristic examples are shown.

The form of the entropy (Eq. 4) follows from the microscopic view of entropy as a measure of uncertainty [43]. For a discrete distribution  $P(y)$  over the variable  $y$ , the functional

$$H[P(y)] = - \sum_y P(y) \ln P(y) \quad (6)$$

is a statistical measure of uncertainty [43, 44]. This quantity is maximised if  $P(y)$  is uniformly spread, and zero if  $P(y)$  is only non-zero in one state. The expression for  $S_\sigma^{\text{eq}}$  in Eq. 4 is a generalisation of Eq. 6 to continuous variables, with the constant  $k_B$  introduced to connect to physical quantities. The postulate that the quantity defined in Eq. 4 is equal to the thermodynamic entropy is the fundamental link connecting statistical mechanics to macroscopic thermodynamics.

The thermodynamic quantities listed above are the natural quantities of interest when we consider a system being manipulated from the outside, and thereby transitioning between two distinct equilibria. However, in molecular systems, we typically set up the system in a non-equilibrium state, and allow it to evolve without further perturbation. We are then generally interested in questions such as: what are the molecular abundances in the eventual equilibrium state, and how fast does the system get there (if at all)? To answer these questions, it is helpful to define *biochemical macrostates*.

## B. Macrostates

The Boltzmann distribution specifies the relative abundances of microstates in equilibrium, and simulation of detailed models typically aims to sample from this distribution. However, microstates are inaccessible in experiment, and inconvenient for theory. Instead, we typically consider biochemical *macrostates*. Rather than keep track of all of the atoms in a DNA molecule or protein, we might simply predict theoretically, measure experimentally or infer from simulation the behaviour of the

position of the centre of mass, or gross conformational features. In other words, we group sets of microstates together into a relatively small number of macrostates, as illustrated schematically in Fig. 1.

In equilibrium, the occupancy of macrostate  $i$  is obtained by integrating over all microstates within it

$$P_\sigma^{\text{eq}}(i) = \frac{Z_\sigma(i)}{Z_\sigma} = \int_{(\mathbf{x}, \mathbf{p}) \in i} d\mathbf{x} d\mathbf{p} \rho \frac{\exp(-E_\sigma(\mathbf{x}, \mathbf{p})/k_B T)}{Z_\sigma}, \quad (7)$$

where I have defined the partial partition function  $Z(i)$ . Typically, we quantify this probability  $P_\sigma^{\text{eq}}(i)$  using the free energy

$$F_\sigma(i) - F_\sigma^{\text{eq}} = -k_B T \ln(P_\sigma^{\text{eq}}(i)) = -k_B T \ln(Z_\sigma(i)/Z_\sigma). \quad (8)$$

Macrostates with high free energy are improbable, and macrostates with low free energy are probable.

In principle, any division into macrostates is valid, although only well-chosen macrostates are helpful. Typically, well-chosen macrostates are either directly identifiable in experiment, amenable to theoretical modelling, or both. Examples might include macrostates labelled by the number of proteins in dimeric complexes; the end-to-end extension of a biomolecule under stress; or the number of base pairs in a DNA hairpin.

There are two conceptually distinct contributions to the free energy  $F_\sigma(i)$ : the average energy of macrostate  $i$ ,

$$U_\sigma(i) = \int_{(\mathbf{x}, \mathbf{p}) \in i} d\mathbf{x} d\mathbf{p} E_\sigma(\mathbf{x}, \mathbf{p}) P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}|i), \quad (9)$$

and the entropy of macrostate  $i$ ,

$$S_\sigma(i) = -k_B \int_{(\mathbf{x}, \mathbf{p}) \in i} d\mathbf{x} d\mathbf{p} P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}|i) \ln(P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}|i)/\rho). \quad (10)$$

Here  $P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p}|i) = P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p})/P_\sigma^{\text{eq}}(i)$  is the equilibrium probability density of occupying microstate  $(\mathbf{x}, \mathbf{p})$  within  $i$ , given that the system is in one of the microstates within macrostate  $i$ . A low average energy implies that a macrostate  $i$  more probable, since individual microstates with lower energies are more probable. A high entropy  $S_\sigma(i)$  implies that many microstates  $(\mathbf{x}, \mathbf{p})$  contribute to  $i$ ; for a given average energy, a macrostate with more accessible microstates is more probable.

## III. FREE ENERGIES OF BIOCHEMICAL REACTIONS

In this section, I will introduce a statistical mechanical approach to biochemical reactions in dilute solution. In the subsequent sections, this basic framework will be applied to a range of contexts of relevance to natural and engineered systems. Before proceeding, it is worth noting that, as in Section II A, I am considering a system  $\sigma$  that can exchange heat with its environment  $\Sigma$ , but

which I have implicitly assumed to occupy a fixed volume  $V_\sigma$ . In chemical contexts, it is often more natural to consider a system maintained at constant pressure  $p$ . In this case,  $V_\sigma$  has to shrink or grow in response to reactions that tend to decrease or increase the internal pressure, respectively. In this case, the Gibbs free energy  $G$ , which plays a similar role to the Helmholtz free energy  $F$  in a constant pressure setting, is the key quantity.

Biomolecular processes, however, occur in aqueous solution, and the enormous numbers of water molecules present dominate the pressure exerted by the system [22]. Reactions between the relatively small number of solute molecules therefore have almost no effect on the pressure and both theoretical work and experimental analysis usually assume constant volume (which is much easier to work with). Nonetheless, free energies of solute states are typically quoted in terms of the Gibbs free energy  $G$ , and the enthalpy  $H$  replaces the average internal energy  $U$ . For internal consistency, I will continue to use  $F$  and  $U$ , but readers familiar with  $G$  and  $H$  should treat them as essentially equivalent. It is also worth noting that in biochemistry, it is more common to use the molar gas constant  $R$  rather than Boltzmann's constant  $k_B$ . This is simply a question of measurement units; if  $R$  is used, all entropies and energies must be given per mole of substance, rather than per particle.

The starting point for our analysis is to treat the solvent implicitly. Formally, this corresponds to integrating over the solvent degrees of freedom in the partition function (Eq. 2), leaving only effective interactions between solute degrees of freedom. In practice, we often just assume that this can be done, and take the effective solute interactions as an input. We then assume that all solutes, or complexes of solutes, can be assigned to a discrete set of molecular species. These species might include ATP, ADP and inorganic phosphate, or a set of individual DNA strands and their complexes: for example, DNA strand  $A$ , DNA strand  $B$  and duplex  $AB$ . It is helpful to define macrostates  $\{N\}$  of the entire solution in terms of the abundances of each of these species,  $\{N\} = (N_A, N_B, \dots)$ . A typical macrostate of a small system, at this level of description is schematically illustrated in Fig. 2.

For each species  $j$ , we can define the single-molecule partition function in the volume  $V_\sigma$ ,  $z_\sigma^j$ . This quantity is analogous to the partition function in Eq. 2, but the integral is performed only over the degrees of freedom of a single solute molecule of species  $j$  in a volume  $V_\sigma$ , with the solvent again treated implicitly. It should be noted that  $z_\sigma^j$  is generally strongly temperature-dependent. In a dilute solution, the overall partition function of a macrostate  $\{N\}$  is essentially given by a product of the individual partition functions, since interactions between molecules that are not in a complex are weak. Hence the degrees of freedom for separate complexes are essentially

independent and the partition function factorises. Thus

$$Z_\sigma(\{N\}) = \prod_j \frac{(z_\sigma^j)^{N_j}}{N_j!}, \quad (11)$$

where the product  $j$  runs over all species types, including complexes. The extra factorial term corrects for overcounting of states that should actually be viewed as indistinguishable, because they are related by the exchange of identical molecules [6, 7, 38]. From this partition function, the free energies of chemical macrostates follow

$$\begin{aligned} F_\sigma(\{N\}) &= -k_B T \ln Z_\sigma(\{N\}) \\ &= \sum_j F_\sigma^j(N_j) \\ &= -k_B T \sum_j \ln \left( \frac{(z_\sigma^j)^{N_j}}{N_j!} \right) \\ &\approx -k_B T \sum_j (N_j \ln z_\sigma^j - N_j \ln N_j + N_j). \end{aligned} \quad (12)$$

Here, I have highlighted the fact that the free energy decomposes into a sum over the contributions from each species,  $F_\sigma^j(N_j)$ , which is a result of the diluteness approximation. The final line uses Stirling's approximation of  $\ln N! \approx N \ln N - N$ , which is highly accurate for large  $N$ . We can thus easily calculate the chemical potential  $\mu_\sigma^j$ :

$$\mu_\sigma^j = \frac{\partial F_\sigma(\{N\})}{\partial N_j} = -k_B T \ln(z_\sigma^j) + k_B T \ln(N_j), \quad (13)$$

In the limit that  $N_j$  is large,  $\mu_\sigma^j$  is simply the difference in free energy arising from adding one molecule of species  $j$  to the system.

Since each  $z_\sigma^j$  is a partition function for a single molecule of species  $j$  in volume  $V_\sigma$ , it will grow proportionally to  $V_\sigma$ ; doubling the volume doubles the number of accessible positions for the molecule, but changes nothing else. It is thus convenient to normalise using a standard volume  $V_0$ , typically taken as the volume in which a single molecule would constitute a concentration  $C_0 = 1/V_0$  of 1 mole per litre. Thus

$$\mu_\sigma^j = -k_B T \ln \left( z_0^j \right) + k_B T \ln \frac{C_j}{C_0}, \quad (14)$$

with a  $z_0^j = z_\sigma^j V_0 / V_\sigma$  dependent on the choice of standard volume  $V_0$ , rather than system volume  $V_\sigma$ . This decomposition separates the chemical potential  $\mu_j$  (or free energy per molecule of species  $j$ ) into a term that depends only on the details of the effective interactions within the species, and a concentration-dependent term [22, 38].

The chemical potential appearing in Eq. 14 is of enormous use in analysing the thermodynamics of molecular systems. In particular, it enables us to calculate the difference in free energy between initial and final macrostates after any given molecular reaction, telling

us how much more likely the final state is than the initial state in equilibrium. To see this, note that every possible reaction  $k$  is associated with a stoichiometric vector  $\nu_{kj}$ , the number of molecules of species  $j$  produced by the reaction  $k$  (this number is negative if species  $j$  is consumed by the reaction). For example, if reaction  $k$  is  $A + B \rightarrow AB$ ,  $\nu_{kA} = -1$ ,  $\nu_{kB} = -1$  and  $\nu_{kAB} = 1$ . Thus the difference in macrostate free energy due to reaction  $k$  is simply [22]

$$\Delta_k F_\sigma \approx \sum_j \nu_{kj} \frac{\partial F_\sigma(\{N_j\})}{\partial N_j} = \sum_j \nu_{kj} \mu_\sigma^j, \quad (15)$$

where the approximation is highly accurate if  $N_j$  is large. Note that  $\Delta_k F_\sigma$ , like  $z_\sigma^j$  and  $z_0^j$ , is generally strongly temperature-dependent. I will continue to specify the reaction in question by a subscript on  $\Delta$ , so that  $\Delta_k$  indicates the change due to reaction  $k$ .

Equally,  $\Delta_k F_\sigma$ , through  $\mu_\sigma^j$ , is dependent on the concentrations of the reactants and products. It is common to consider the value of  $\Delta_k F_\sigma$  at the reference concentration  $C_j = C_0 = 1$  M,  $\Delta_k F^0$ , as a “standard” free energy of the reaction.  $\Delta_k F^0$  can also be further subdivided into standard energetic and entropic contributions,  $\Delta_k U^0$  and  $\Delta_k S^0$ . This is convenient for book-keeping purposes, but it should be noted that this “standard” value is dependent on the (arbitrary) choice of  $C_0$ , unless the number of reactants and products is the same. For example, for the simple bimolecular binding reaction  $A + B \rightarrow AB$ ,

$$\Delta_{A+B} F^0 = -k_B T \ln \frac{z_0^{AB}}{z_0^A z_0^B}, \quad (16)$$

and the term inside the logarithm scales as  $V_0$ . In such cases, it is inadvisable to over-interpret the sign and magnitude of the standard free energy; whether it is positive or negative depends upon the arbitrary standard concentration, which is 1 M in general. This dependence on  $V_0$  feeds through to the standard entropy  $\Delta_k S^0$ , so it is also unwise to read too much into the sign and magnitude of this quantity when the numbers of reactants and products differ. Indeed, for macromolecules such as DNA, RNA and proteins, 1 M corresponds to an incredibly concentrated solution where the dilute approximations above break down, and behaviour such as the formation of liquid crystals is observed [45, 46]. For processes involving unequal numbers of reactants and products, therefore, the standard free energy and entropy exist purely as book-keeping devices, and never describe the actual properties of a reaction. In typical dilute molecular systems, the concentration of the relevant components is orders of magnitude lower. Thus actual values of  $\Delta_k F_\sigma$  and  $\Delta_k S_\sigma$  are significantly more positive than the standard values for assembly reactions in which the number of reactants exceed the number of products.

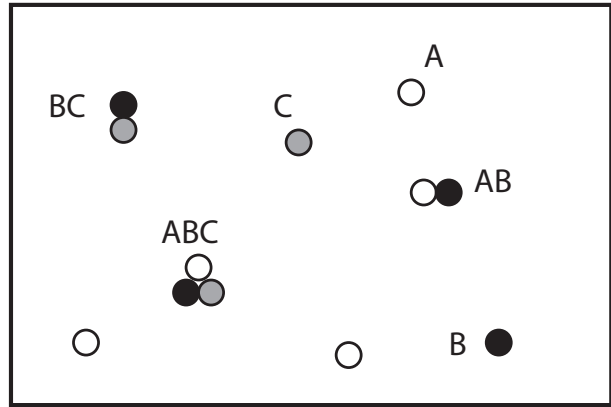


FIG. 2. A self-assembling system, in which monomers  $A$ ,  $B$  and  $C$  combine to form a trimer  $ABC$ . The macrostate in this snapshot is  $N_A = 3$ ,  $N_B = 1$ ,  $N_C = 1$ ,  $N_{AB} = 1$ ,  $N_{BC} = 1$ ,  $N_{AC} = 0$  and  $N_{ABC} = 1$ .

#### IV. THE APPLICATION OF EQUILIBRIUM THERMODYNAMICS TO THE DESIGN OF SELF-ASSEMBLING SYSTEMS

##### A. Self-assembly

Self-assembly occurs when molecules are mixed and autonomously bind to produce non-trivial structures. It should be distinguished from step-by-step directed synthesis in which each stage is separately coordinated by an experimenter through careful manipulation of solution conditions [47], although temperature ramps are often used to optimise results, as analysed in some detail in Ref. [48]. In biology, active protein complexes [49] and virus capsids [50, 51] must assemble accurately from their components to be functional; the rise of nucleic acid nanotechnology has facilitated the design of artificial systems that can mimic this behaviour, allowing precisely-controlled finite-size nanostructures [29, 31, 52, 53].

The simplest design strategy is to ensure that microstates with many well-assembled structures are common in equilibrium. A mixture of components should then tend towards a high yield of target structures given enough time. In this section, we shall consider the thermodynamics of equilibrium states of self-assembling systems, and the implications for system design. To provide context, I will apply the general results to the challenge of assembling three molecular building blocks  $A$ ,  $B$  and  $C$  into the complex  $ABC$ , as illustrated in Fig. 2. Molecules  $A$ ,  $B$  and  $C$  could each be DNA strands or proteins, for example.

We shall work within the formalism of Section III, considering the probability of observing macrostates of molecular abundances in equilibrium. The most likely macrostate  $\{N\}$  is the one that maximises  $Z_\sigma(\{N\})$ , or equivalently minimises  $F_\sigma(\{N\})$ , subject to the con-

straints of stoichiometry. If there are a large number of molecules, as in typical experiments, fluctuations about the most likely macrostate are relatively small in equilibrium [6, 7, 38, 43], and hence we can infer equilibrium properties purely by analysing this most likely macrostate.

### B. Identifying the typical state of a self-assembling system in equilibrium

To minimise  $F_\sigma(\{N\})$ , we must find  $\{N\}$  such that no possible change of  $\{N\}$  due to a reaction could reduce  $F_\sigma(\{N\})$ . In the limit of many molecules, this is equivalent to identifying the  $\{N\}$  for which  $F_\sigma(\{N\}) - F_\sigma(\{N'\}) = 0$  for every possible reaction  $\{N\} \rightarrow \{N'\}$ . From Eq. 15 we therefore require

$$\Delta_k F_\sigma = \sum_j \nu_{kj} \mu_\sigma^j = 0, \quad (17)$$

Thus the typical equilibrium state is the one in which chemical potentials are balanced for all reactions – a fundamental result [22, 38].

Substituting our expression for  $\mu_\sigma^j$  (Eq. 14) into Eq. 17, we immediately see that at equilibrium

$$\prod_j \left( \frac{C_j}{C_0} \right)^{\nu_{kj}} = \prod_j \left( z_0^j \right)^{\nu_{kj}}, \quad (18)$$

for all reactions  $k$ . In the case of  $A + B \rightarrow AB$ ,

$$\frac{C_{AB}}{C_A C_B} = \frac{1}{C_0} \frac{z_0^{AB}}{z_0^A z_0^B} = K_{A+B}^{\text{eq}}, \quad (19)$$

in which I have introduced the equilibrium constant  $K_{A+B}^{\text{eq}}$ . The quantity  $K_{A+B}^{\text{eq}}$  is known as a constant because it depends only on the details of the interactions within each species as represented through  $z_0^j$ ; it is independent of system volume or the number of molecules present, although it will depend on quantities such as the temperature. It is also independent of the arbitrary reference volume  $V_0$ , since each  $z_0^j$  scales with this volume. The result is ubiquitous in physical chemistry, and immediately generalises for other reactions [22].

If  $K^{\text{eq}}$  is known from earlier experiments for each reaction  $k$ , or can be predicted from underlying theory, then variants of Eq. 19 can be constructed for each possible reaction and the typical concentrations in equilibrium can be found by solving the resultant simultaneous equations (when augmented with any conservation laws). Note that a relation such as Eq. 19 exists for every reaction at equilibrium, regardless of whether the reactants are involved in other reactions. Additionally, the same procedure can be followed for not just a single reaction  $k$ , but a series of reactions. For example, the combined reactions  $A + B \rightarrow AB$  and  $AB + C \rightarrow ABC$  imply that the relationship

$$\frac{C_{ABC}}{C_A C_B C_C} = \frac{1}{C_0^2} \frac{z_0^{ABC}}{z_0^A z_0^B z_0^C} = K_{A+B+C}^{\text{eq}}, \quad (20)$$

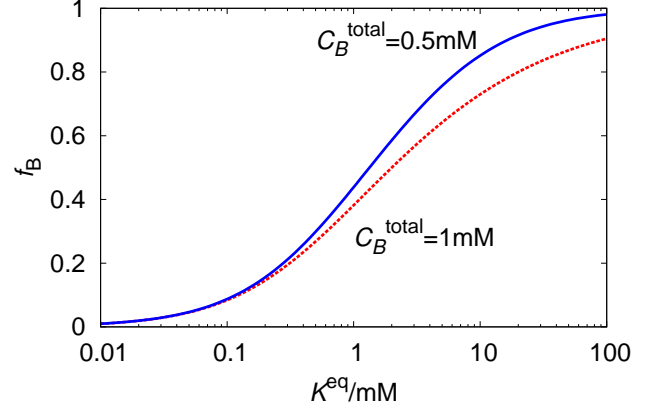


FIG. 3. Non-stoichiometric conditions can increase the fractional of certain molecules incorporated into target complexes. I plot the equilibrium fraction of  $B$  molecules incorporated into  $AB$  ( $f_B$ ) by the reaction  $A + B \rightleftharpoons AB$ , as a function of  $K_{A+B}^{\text{eq}}$ . stoichiometric mixture. Yield curves are plotted for  $C_B^{\text{total}} = C_A^{\text{total}} = 1 \text{ mM}$ , and  $A$ :  $2C_B^{\text{total}} = C_A^{\text{total}} = 1 \text{ mM}$ .

is meaningful even if there is no direct  $A + B + C \rightarrow ABC$  reaction (which is essentially impossible without the formation of intermediate complexes).

To design a system that can self-assemble efficiently into  $ABC$ , therefore, we should choose molecules that provide a large  $K_{A+B+C}^{\text{eq}}$ , and also large  $K_{A+BC}^{\text{eq}}$ ,  $K_{AB+C}^{\text{eq}}$  and  $K_{AC+B}^{\text{eq}}$ , since  $ABC$  needs to out-compete other potential species. Similarly, low values of  $K^{\text{eq}}$  for reactions that produce off-target species such as  $AABB$  are important in preventing mis-assembly.

An often under-appreciated fact is that total concentrations are important; it is not uncommon to hear a melting temperature – approximately the point at which 50% of the maximum possible yield has been reached – quoted without reference to component concentrations. However, the fractional yield in equilibrium is sensitive to the initial concentrations of reactants. Lower total concentrations of  $A$ ,  $B$  and  $C$  imply a larger reduction in the concentration of complexes in equilibrium, since the numerator in relationships such as Eq. 20 must change enough to compensate for the reduction of all of the concentrations in the denominator.

Depending on the context, it may be advantageous to use non-stoichiometric mixtures of components – for example, an excess of  $A$  and  $B$  relative to  $C$  when assembling  $ABC$ . Doing so significantly enhances the fraction of  $C$  molecules incorporated into  $ABC$  structures in equilibrium, at the expense of leaving a pool of  $A$  and  $B$  which cannot possibly contribute to a target. To illustrate this effect, I plot the fractional yield  $C_{AB}/C_B^{\text{total}}$  in the reaction  $A + B \rightleftharpoons AB$  as a function of  $K_{A+B}^{\text{eq}}$  in Fig. 3. I consider two cases, one with a stoichiometric mixture  $C_B^{\text{total}} = C_A^{\text{total}} = 1 \text{ mM}$ , and one with an excess of  $A$ :  $2C_B^{\text{total}} = C_A^{\text{total}} = 1 \text{ mM}$ . It is clear that the fraction of  $B$  molecules incorporated into complexes tends



to unity much more quickly. This approach is taken in the construction of scaffolded DNA origami [29]; an excess of staples is added to enhance the fraction of scaffold strands incorporated into well-formed structures. It is a particularly useful approach if it is free strands of one type or the other might cause leak reactions. Murugan *et al.* have also proposed that concentrations of reactants could be judiciously chosen to avoid the formation of undesired off-pathway structures [54].

Finally, it is worth noting that the equilibrium constants  $K^{\text{eq}}$  are determined exclusively by the properties of the reactant and product species, through the ratio of the appropriate partition functions. Thus relationships such as Eqs. 19 and 20 *cannot* be altered by other molecules that are not produced or consumed by the process, or by the properties of intermediate complexes. An change in the equilibrium constant, as measured through the relative concentrations in equilibrium, can only arise from a change in the biochemistry of the initial and final species. This fact will be important when I discuss the role of non-equilibrium catalysts in Section VII.

### C. The meaning of the equilibrium constant, and estimating its value

Relationships between concentrations such as Eqs. 19 and 20 can also be derived more by assuming that reactions obey mass-action kinetics: *i.e.*, that rates are proportional to concentrations of reactants involved. However, the statistical mechanical approach has two important advantages.

Firstly, the fact that relationships such as Eqs. 19 and 20 hold even without a direct one-step reaction linking reactants and products, and regardless of whether the species are involved in other reactions, is made clear from this thermodynamic perspective. But perhaps even more importantly, the derivation presented highlights the physical meaning of the equilibrium constant  $K^{\text{eq}}$ ; it is determined by the ratios of partition functions of the molecular species involved in the reaction. The quantities  $z_0^j$  are partition functions for complex  $j$  in volume  $V_0$ ; species with favourable (low) internal energies, or many accessible states, are favoured.

Crucially, these quantities  $z_0^j$  (or at least the relevant ratios) can be predicted theoretically. For complex models [39–42], ratios of partition functions are the natural quantities to extract from simulation [55, 56], enabling direct prediction of equilibrium constants and hence comparison to experiment. Simpler approaches such as the nearest-neighbour models of DNA and RNA thermodynamics use basic postulates about  $z_0^j$  to predict analytically the ratios of partition functions, and hence equilibrium constants, of an enormous number of self-assembling systems using a small set of parameters [57, 58]. These widely-used tools have been a fundamental component of the growth of nucleic acid nanotechnology and have facilitated the analysis of natural RNA

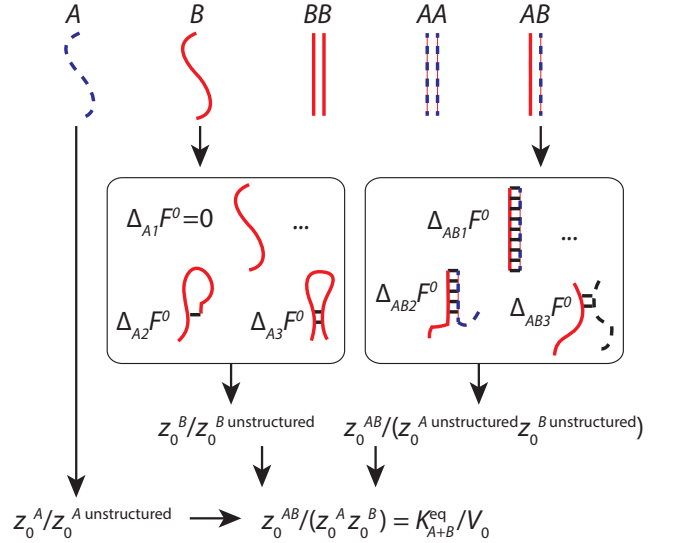


FIG. 4. A schematic illustration of the calculation of equilibrium constants by algorithms such as Nupack [60], based on nearest-neighbour models of nucleic acid thermodynamics [57, 58]. Firstly, all possible complexes up to a maximum size (in this case 2) are enumerated. For each of these complexes, and the individual strands, the possible-base pairing macrostates are identified and assigned a free energy relative to the unstructured state. Summing over all contributions for one complex gives the partition function of the complex relative to the unstructured constituent strands; combining this with similar quantities estimated for the single strands themselves gives  $K_{A+B}^{\text{eq}}$ .

circuitry in cells, for example in Ref. [59].

It is hard to overestimate the usefulness of such a predictive tool, even given its finite accuracy. Without it, the systematic design of complex nucleic acid circuits from scratch would be far more challenging – particularly in terms of eliminating unintended interactions – and would require the measurement of many equilibrium constants  $K^{\text{eq}}$  for each design.

It is instructive to consider how the nearest-neighbour model can be used by utilities such as Nupack [60] to predict the concentration of complexes formed in equilibrium after mixing nucleic acids. Firstly, given a set of strands, all possible complexes below a certain size can be identified (Fig. 4). For each of these complexes  $j$ , it is then necessary to estimate the equilibrium constant  $K_j^{\text{eq}}$  for complex formation from the constituent single strands at the appropriate temperature – with these equilibrium constants, the task reduces to solving a set of simultaneous equations involving expressions such as Eqs. 19 and 20 and conservation laws.

Equilibrium constants are estimated by working at the level of macrostates defined by the pattern of base pairing – a finer resolution than simply identifying the complexes present, but still far from a true microscopic enumeration of microstates. To estimate  $K_{A+B}^{\text{eq}}$ , for example, all base-pairing macrostates of both the complex  $AB$  and

the individual strands  $A$  and  $B$  are enumerated (Fig. 4). For each macrostate  $i$ , the nearest-neighbour model predicts the standard free energy relative to a completely unstructured (base-pair free) macrostate using a small set of universal parameters that depend on the sequence of base-pair steps, and the context at the end of the continuous base-paired stations. From these free energies, the relative partition functions of macrostates can be calculated, and by summing over the partition function contributions from all macrostates for both the complex and the individual strands, the ratio  $z_0^{AB}/(z_0^A z_0^B)$  can be estimated. Hence  $K_{A+B}^{\text{eq}}$  can be predicted through Eq. 19, and used to infer complex concentrations given the total concentrations of all strands.

## V. THERMODYNAMICS AS A BASIS FOR THE DESIGN OF KINETIC MODELS

### A. The importance of kinetics

Designing an equilibrium state to be consistent with a high yield of a self-assembling structure is a typical step taken towards successful self-assembly. Indeed, this is the main strategy employed hitherto in the field of nucleic acid nanotechnology when designing structures. However, the existence of a high yield in equilibrium doesn't guarantee successful assembly in finite time – the system might become trapped in metastable states, and fail to approach the equilibrium yield over a reasonable timescale. To understand why, it is important to develop kinetic models that can explore dynamical trajectories taken by systems.

Static self-assembled structures are also not the only possible type of molecular system. There has been recent interest in non-equilibrium or dissipative self-assembly [61], in which assembled structures are maintained in a non-equilibrium rather than an equilibrium steady state by a continuous input of energy. Indeed, this is almost a minimal description of a living organism. On a more detailed level, natural molecular circuits generate motion [22, 62], act as oscillatory clocks [63], and dynamically sense their environment [64–67]; researchers are now designing artificial systems with similar functionality [35, 68, 69]. In these dynamical systems, reaction kinetics is inherently important.

### B. Molecular systems as stochastic processes

Statistical mechanics is inherently probabilistic; the Boltzmann distribution (Eq. 1) is, after all, a probability distribution for finding the system  $\sigma$  in a given microstate. It is therefore natural to describe system dynamics using a stochastic (or random) process [70] over these microstates. At a given time  $t$ , the system occupies a microstate  $(\mathbf{x}, \mathbf{p})$  with a probability density  $P_\sigma(\mathbf{x}, \mathbf{p}, t)$ ; system dynamics lead to an evolution of this distribution

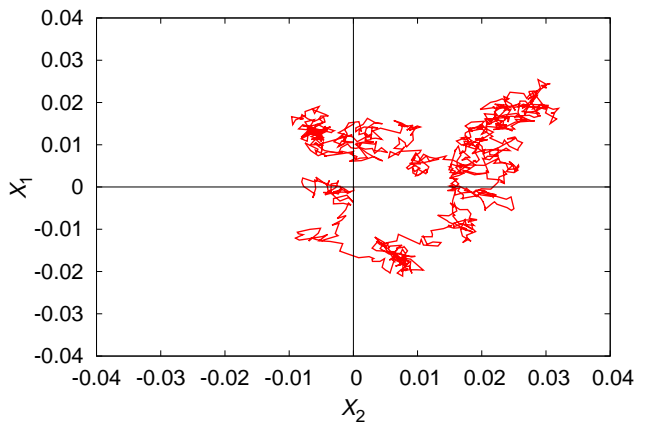


FIG. 5. Coarse-grained dynamics can be non-Markovian, even if the underlying dynamics at the macrostate level is Markovian. Plotted is the trajectory of a simple Markov process (a 2D Ornstein-Uhlenbeck process [70]) over continuous variables  $X_1$  and  $X_2$ . Dividing the state space into macrostates according to the sign of  $X_1$  and  $X_2$  leads to highly non-Markovian dynamics at the macrostate level; destinations and dwell times do not exhibit the memoryless property.

over time. After a long time,  $t \rightarrow \infty$ , and in the absence of external driving, the system should relax towards a stationary (time-independent) distribution given by the Boltzmann distribution  $P_\sigma^{\text{eq}}(\mathbf{x}, \mathbf{p})$  of Eq. 1.

For an ideal memoryless environment, the stochastic process should be Markovian [70]; *i.e.*, the future evolution of  $P_\sigma(\mathbf{x}, \mathbf{p}, t)$  only depends on the past via the current value of  $P_\sigma(\mathbf{x}, \mathbf{p}, t)$ . Equivalently, the outcome of leaving a microstate  $(\mathbf{x}, \mathbf{p})$  is independent of the route by which the system reached  $(\mathbf{x}, \mathbf{p})$ . Of course, when analysing molecular systems, we often work at the level of chemical macrostates. We describe the system through the abundances of species, or perhaps the hydrogen-bonding patterns, rather than individual positions and momenta. In this case, our state space is discrete and the stochastic process involves a series of transitions by which the system undergoes discrete hops between macrostates.

Generally, it is assumed that the stochastic process is also Markovian at the macrostate level [10, 71, 72]. In other words, transitions from macrostate  $i$  to macrostate  $j$  occur at a fixed rate  $k_{ji}$  (frequently zero), meaning that the time spent in state  $i$  prior to a transition is exponentially distributed; the average lifetime is given by  $\tau_i = 1/\sum_j k_{ji}$ ; and the destination state is independent of the state from which the system entered  $i$  and the time spent within  $i$ . Formally, given this assumption, the probability distribution evolves according to the familiar Master equation [70]. This type of model is so universal that it is often not realised just how big an assumption it is to treat coarse-grained dynamics as Markovian.

Consider Fig. 5, in which I have divided a two-dimensional microstate space into macrostates in an essentially arbitrary fashion, and plotted a sample trajec-



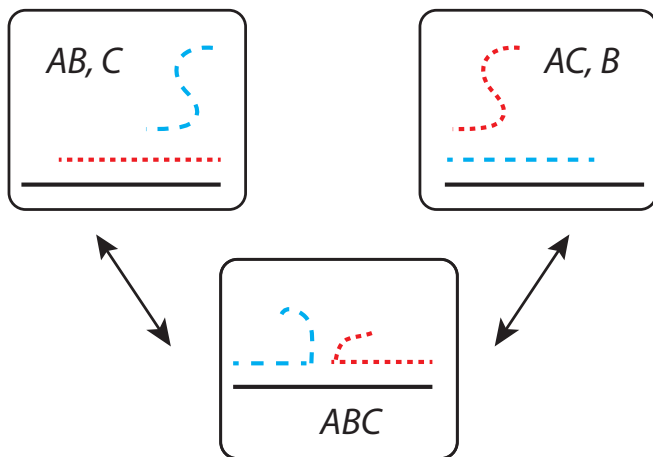


FIG. 6. A practically-relevant example of a coarse-graining into macrostates that is inappropriate for treatment with Markovian dynamics. Consider a DNA-based toehold-exchange reaction, in which  $B$  and  $C$  compete for binding to  $A$  via strand displacement. It is perfectly possible to define three macrostates as shown above, and indeed doing so would enable the calculation of molecular abundances in equilibrium. However, it is inappropriate to model the dynamics as Markovian, because the microstates of the  $ABC$  complex are not typically fully-explored before it dissociates.

tory of a simple process that is Markovian at the microstate level. Fundamentally, even if the dynamics is Markovian at the microstate level, coarse-graining introduces memory into the process. Trajectories that enter macrostate  $i$  from macrostate  $j$  are close to the border between the two, and hence are likely to quickly cross back into  $j$ . As a result, both the transition time *and* transition destination should, in general, depend in a complex manner the previous macrostates visited by the system, and the time for which the current macrostate has been occupied.

For a more concrete example, consider the strand exchange reaction shown in Fig. 6. Strand exchange is a basic process underlying much of DNA computation [34, 73, 74]. It might be tempting to describe the system using three macrostates:  $A$  bound to  $B$  only; an  $ABC$  complex; and  $A$  bound to  $C$  only. Indeed, such an approach is sensible if one is only interested in the relative abundances of complexes in equilibrium. However, the system dynamics cannot be well-described by a Markov process at this level. The need to initiate and complete branch migration to exchange base pairs between  $AB$  and  $AC$  duplexes means that an  $ABC$  complex formed by  $AB + C \rightarrow ABC$  is in reality much more likely to dissociate into  $AB + C$  than a complex formed by  $AC + B \rightarrow ABC$ , violating the assumptions of a Markov process. Even splitting the  $ABC$  macrostate into two separate macrostates, depending on whether  $AB$  or  $AC$  contains the most base pairs [73], does not provide a satisfactory treatment of the system; it is necessary to resolve macrostates on at least the base-pair level to

provide a predictive Markov model of system dynamics [75]. A similar example in cells is the translation of RNA [76]; if “ribosome bound to RNA” is treated as a single macrostate, rather than modelling codon incorporation as individual steps, a highly unrealistic exponential distribution of times for translation will be obtained.

So when is a Markov assumption reasonable? This is a subtle problem, but the basic idea is that macrostates must be carefully chosen so that a transition involves passing through an unfavourable (high energy or low entropy) set of microstates around the boundary. In this case, the system typically spends a long time fully exploring each macrostate before a sudden hop to a neighbouring one. Transitions, which are complete when the system has fully crossed the unfavourable microstates on the boundary, are rare – the time taken waiting to see a transition is long compared to the duration of the transition. When these assumptions hold, any memory of the previous macrostate is lost whilst exploring the new one, allowing the Markov assumption [10, 71, 72].

A simple example is given by comparing the trajectories of particles in two potential energy wells. In the first case, the well is quadratic, with a single minimum at  $x = 0$ ; in the second, it is quartic, with two minima at  $x = -1$  and  $x = +1$  (see Fig. 7(a)). In both cases it is formally possible to define macrostates according to whether the particle occupies  $x < 0$  or  $x > 0$ . However, only in the first case, in which the transition from  $x < 0$  to  $x > 0$  is associated with climbing over an unfavourable energy barrier, is it reasonable to describe this process using a Markov model at the macrostate level (compare the two trajectories in Fig. 7(b)). Returning to the toehold exchange reaction in Fig. 6, the three-state Markov model fails because a system that enters the  $ABC$  macrostate isn’t likely to fully explore that macrostate prior to leaving it. Since branch migration is slow [75], the strand that has just bound will often detach before all branch migration intermediates have been explored, invalidating the requirements for a Markov model at the level of these macrostates.

From this point onwards I will assume a biomolecular system with discrete macrostates that have been well-chosen. Thus the dynamics is Markovian and can be well described by a master equation with rate parameters  $k_{ji}$  [70]. What I will say will also be applicable to a full description at the level of microstates ( $\mathbf{x}, \mathbf{p}$ ).

### C. Detailed balance

The transition rates  $k_{ji}$  between all pairs of macrostates fully define system behaviour, given a particular initial condition. Knowledge of system thermodynamics (*ie.*, the free energy of macrostates  $F_\sigma(i)$ ) doesn’t specify kinetics, but it does place strong and important restrictions. Firstly as noted in Section VB, the system should eventually relax to the equilibrium (Boltzmann) distribution over macrostates. Thus knowledge of the



than they would from  $j$  to  $i$ . In equilibrium, this tendency is compensated for by the relative population size, yielding detailed balance. This is true even if it takes several steps to reach  $j$  from  $i$ , since Eq. 22 holds for each of those steps. If  $(F_\sigma(i) - F_\sigma(j))/k_B T \gg 1$ , then we expect to see  $i \rightarrow j$  occur spontaneously during trajectories, but we will essentially never observe a system starting in  $j$  and transitioning to  $i$  (unless we force it from the outside). In the context of a chemical reaction, for example  $A + B \rightleftharpoons AB$ , we expect to see systematically more transitions from left to right whilst  $\Delta_{A+B} F_\sigma = \mu_\sigma^{AB} - (\mu_\sigma^A + \mu_\sigma^B) > 0$ . On a large scale with many molecules, we will effectively see the reaction spontaneously proceed in one direction determined by the sign of  $\Delta_{A+B} F_\sigma$ , until equilibrium ( $\Delta_{A+B} F_\sigma = 0$ ) is reached.

#### D. Designing a kinetic model for molecular systems

We are now in a position to build kinetic models for molecular systems that are thermodynamically sound. Doing so is essential in developing an interpretable model of self-assembly, but it is also relevant to other molecular contexts, as I will discuss in Sections VIA, VII and VIII. Sometimes, it is convenient to model the stochastic dynamics explicitly, perhaps by considering only a small number of molecules. An alternative is to assume a large number of molecules, and consider the deterministic evolution of the average concentrations. In either case, the key ingredients in the model are the transition rates  $k_{ij}$ . It is typical to assume mass-action kinetics to handle the concentration-dependence of reactions (rates per unit volume are proportional to concentrations of reactants). For example, in the reaction  $A + B \rightarrow AB$ , the macrostate-dependent rate for the binding transition  $k_{N_{AB}+1, N_{AB}}(\mathcal{C}_A, \mathcal{C}_B)$  is given by

$$k_{N_{AB}+1, N_{AB}}(\mathcal{C}_A, \mathcal{C}_B) = \phi_{\text{bind}} \mathcal{C}_A \mathcal{C}_B V_\sigma, \quad (23)$$

where  $\phi_{\text{bind}}$  is a bimolecular rate constant, and the reverse rate is given by

$$k_{N_{AB}, N_{AB}+1}(\mathcal{C}_B) = \phi_{\text{unbind}} \mathcal{C}_B V_\sigma, \quad (24)$$

where  $\phi_{\text{unbind}}$  is a unimolecular rate constant for unbinding. Thus, using Eq. 14 and Eq. 16,

$$\frac{\phi_{\text{bind}}}{\phi_{\text{unbind}}} = V_0 \exp(-\Delta_{A+B} F^0 / k_B T), \quad (25)$$

hopefully a familiar result.

It is a near-impossible task to write down a set of rates or rate constants that obey detailed balance directly, for anything but the most simple of systems. Instead, it is advisable to start from a model for the free energies  $F_\sigma(i)$ , and then impose detailed balance using Eq. 22, or variants of Eq. 25, for each pair of reactions. Thus the relative rates for each pair of transitions are given, and

all that remains is to specify the absolute rates of one of them, which again requires some physical model. As an example, we might consider a small DNA nanostructure, whose assembly involves binding and unbinding of duplex sections. A simple assumption might be that all binding transitions have the same bimolecular rate constant (perhaps  $\sim 10^6 \text{ Ms}^{-1}$ ), which sets  $k_{ij}$  for all binding transitions. The inverse  $k_{ji}$  transition rates then follow from the free energy of binding,  $F_\sigma(i) - F_\sigma(j)$ , which might be estimated via the nearest-neighbour model.

In general, both the problem of estimating  $F_\sigma(i)$  and the absolute rates of one of each pair of transitions can be subtle [75, 77–79], and the consequences for the dynamics can be profound. In particular, it is not always obvious how a change in  $\Delta_k F_\sigma$  might be manifest in the rates. For example, consider the association of two DNA strands  $A$  and  $B$  that can form hairpins in the single-stranded state. These hairpins serve to make the standard free energy of formation of a duplex  $\Delta_{A+B} F^0$  less favourable (less negative), and hence reduce the ratio  $\phi_{\text{bind}}/\phi_{\text{unbind}}$  through Eq. 25. It might be natural to assume that hairpins essentially reduce the rate of binding  $\phi_{\text{bind}}$ , and indeed hairpins are deliberately used for this purpose to create metastable systems [80–83]. However, experimental evidence suggests that in some cases,  $\phi_{\text{bind}}$  is reduced by much less than  $\exp(-\Delta_{A+B} F^0 / k_B T)$ , and implying that in fact most of the reduction in free energy is manifest as an increase in the off-rate  $\phi_{\text{unbind}}$  [84]. This observation is supported by detailed simulation [85], in which it is observed that hairpins form prior to full dissociation and stabilise the partially-melted state, accelerating unbinding. Similarly, experimental work supported by theory shows that the overall forwards rate of strand displacement and exchange reactions can be adjusted by orders of magnitude at a fixed overall standard free energy of reaction [73, 86].

Despite these subtleties, kinetic models of complex molecular systems can provide deep insight into function. For example, the exquisite data provided by Zhang and Winfree on the rates of DNA strand displacement and toehold exchange reactions [73] allow thermodynamically consistent modelling of those key processes, underlying the systematic design of complex molecular circuits using tools such as DSD [87]. In turn, the physical basis of these parameters have been probed by more detailed modelling [75] at the base-pair level of description. Similarly, it has long been known that kinetics, rooted in a free-energy landscape, is fundamental to understanding how proteins fold [88]. Recent work has explored and manipulated the kinetics of assembly DNA origami and DNA brick assembly in a similar fashion, highlighting the importance of subtle cooperative effects [79, 89–91].

## VI. REVERSIBILITY IN BIOCHEMICAL REACTIONS

### A. Including reverse transitions in modelling

One immediate consequence of Eq. 22 is that if a forwards reaction is possible, so to is its reverse, with relative rates determined by the initial and final free energies. This concept is known as *the principle of microscopic reversibility*, and has far-reaching consequences. This principle suggests that both forwards and backwards transitions should always be explicitly included in any model. However, successful modelling is often done by treating certain reactions as totally irreversible. For example, nobody models transcription of RNA in cells by considering the reverse process by which RNA returns to the DNA and is destroyed base by base whilst in contact the gene that encoded it [92]. Similarly, strand displacement kinetics is often fitted by assuming that the reaction proceeds to 100% completion [86].

In general, whether it is reasonable to neglect reverse transitions depends upon the context, and the purpose of the modelling. Processes such as RNA transcription involve constant input of chemical fuel (this situation will be discussed in Section VII), and are thus far from equilibrium; in such cases, backwards transitions can be rendered irrelevant by the presence of alternative pathways. For example, RNA is digested by exo- and endonucleases in the cell, instead of needing to be destroyed by the reverse of transcription. Unless the modeller is interested in the actual thermodynamic work being done by the fuel in such cases, or in the case of relatively weak driving, explicit modelling of the reverse reactions is unimportant. For strand displacement, with a sufficiently long toehold, the equilibrium state is so biased towards the product of displacement that reverse reactions, leading to a residual concentration of the input, can be neglected. In many other contexts, however, it is important to include microscopically reversible transitions, because the free energy of transitions is relatively weak (as is the case of self-assembly near the melting temperature); because the overall thermodynamics is of interest to the modeller; or because reverse reactions, despite being typically slow, have a profound impact on system behaviour.

For example, in the field of molecular computation and algorithms, it is common to assume that reactions can be made totally irreversible [35, 74, 93–95]. An example is an algorithm for computing the parity (even/odd nature) of an initial number of molecules of type  $A$ , as discussed by Cummings *et al.* [93] and illustrated in Fig. 9 (a). The algorithm introduces two other species,  $B$  and  $C$ , and the (assumed microscopically irreversible) reactions



If the system is initially prepared with  $N_A^0$  molecules of type  $A$ , and 1 molecule of type  $B$ , it can be seen that the

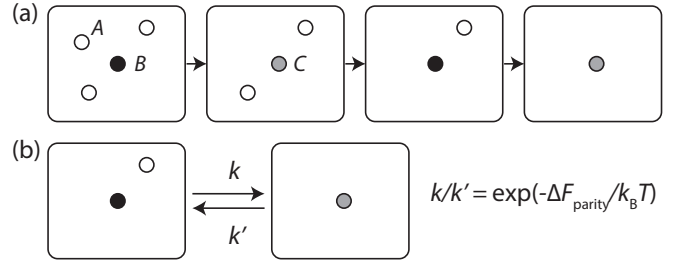


FIG. 9. Schematic illustration of the parity-computing algorithm of Cummings *et al.* [93]. (a) Reactions consume species  $A$ , and switch  $B$  into  $C$  and vice versa. Thus a system initiated with an odd number  $N_A^0$  of  $A$  and a single  $B$  will result in an isolated  $C$  (as shown). Alternatively, a system initiated with an even number  $N_A^0$  of  $A$  molecules will result in a single isolated  $B$ . (b) The consequence of microscopic reversibility is a finite backwards rate for all transitions; I take the overall difference in free energy between  $N_A = 1$  and  $N_A = 0$  to be  $\Delta F_{\text{parity}}$

reactions will interconvert  $B$  and  $C$  (retaining  $N_B + N_C = 1$ ) whilst reducing the number of  $A$  molecules. The final state will be a single molecule of type  $B$  if  $N_A^0$  is even, and a single molecule of type  $C$  if  $N_A^0$  is odd. The output of the network is thus the state of the  $B/C$  molecule in the limit of long time, which reports on the parity of  $N_A^0$ . Note that these reactions can in principle be implemented with DNA using a large supply of implicit ancillary molecules [74, 96].

What happens if we consider reverse reactions, so that  $C \rightarrow A + B$  and  $B \rightarrow A + C$  are also possible, as in Fig. 9 (b)? Then there is a finite chance that a system observed in the long-time limit will contain a  $B/C$  molecule that does not reflect the parity of  $N_A^0$ . Specifically, let us assume that for a single  $A$  and a single  $B$  being converted into a single  $C$  in the volume  $V_\sigma$ , in the absence of all other molecules of type  $A$ ,  $B$  and  $C$ , the free-energy difference between macrostates is  $F_\sigma(N_A = 0, N_B = 0, N_C = 1) - F_\sigma(N_A = 1, N_B = 1, N_C = 0) = \Delta F_{\text{parity}}$  (Fig. 9 (b)). For simplicity, let us also assume that the same free-energy difference applies to  $A + C \rightarrow B$ ,  $F_\sigma(N_A = 0, N_B = 1, N_C = 0) - F_\sigma(N_A = 1, N_B = 0, N_C = 1) = \Delta F_{\text{parity}}$ . These free energies include the contributions from any implicit ancillary molecules. Provided the ancillary molecules are in excess (their concentrations are essentially unaffected by the reactions involving  $A$ ,  $B$  and  $C$ ), then the probability of observing  $N_A$  molecules of type  $A$  in equilibrium is given by

$$\begin{aligned} P_\sigma^{\text{eq}}(N_A = 1) &= P_\sigma^{\text{eq}}(N_A = 0) \exp(-\Delta F_{\text{parity}}/k_B T), \\ P_\sigma^{\text{eq}}(N_A = 2) &= \frac{1}{2} P_\sigma^{\text{eq}}(N_A = 0) \exp(-2\Delta F_{\text{parity}}/k_B T), \\ P_\sigma^{\text{eq}}(N_A = 3) &= \frac{1}{6} P_\sigma^{\text{eq}}(N_A = 0) \exp(-3\Delta F_{\text{parity}}/k_B T), \\ P_\sigma^{\text{eq}}(N_A) &= \frac{1}{N_A!} P_\sigma^{\text{eq}}(N_A = 0) \exp(-N_A \Delta F_{\text{parity}}/k_B T). \end{aligned} \quad (27)$$

Each term includes an additional factor,  $e^{-\Delta F_{\text{parity}}/k_B T}$ , to account for the additional reaction that must take place to reach  $N_A = 0$ . The  $N_A!$  factor accounts for the fact that the free energy difference between states with  $N_A$  and  $N_A - 1$  depends on the number of  $A$  present, as we previously saw in Section III. One way to confirm the exact dependence is to note that the rate for  $C \rightarrow A + B$  and  $B \rightarrow A + C$  should be independent of  $N_A$ , but  $A + B \rightarrow C$  and  $A + C \rightarrow B$  should occur with an overall rate proportional to  $N_A$ . Incorporating this into the free energy gives the factors in Eq. 27 (a logarithmic growth in the free energy difference with  $N_A$ ).

When  $N_A$  is odd, the readout from the reporter molecule  $B/C$  gives an incorrect readout for the parity of  $N_A^0$ . It can be seen that the even and odd terms of Eq. 27 correspond to terms in the expansion of hyperbolic functions of  $e^{(-\Delta F_{\text{parity}}/k_B T)}$ . Thus

$$\frac{P_{\text{correct}}}{P_{\text{incorrect}}} = \frac{P_{\sigma}^{\text{eq}}(N_A \text{ even})}{P_{\sigma}^{\text{eq}}(N_A \text{ odd})} \approx \coth \left( e^{(-\Delta F_{\text{parity}}/k_B T)} \right). \quad (28)$$

For  $\Delta F_{\text{parity}}$  large and negative – when the interconversion of an isolated  $A$  and  $B$  into an isolated  $C$  is favourable – the algorithm is accurate. For lower values of  $\Delta F_{\text{parity}}$ , the accuracy is reduced. Of course, in an abstract design it is possible to imagine  $\Delta F_{\text{parity}}$  is as large as possible, but the need to do this should be noted. It is also worth highlighting the fact that  $\Delta F_{\text{parity}}$  is *not* the standard free energy of the reaction; it is the free energy difference between a macrostate with a single  $A$  and a single  $B$  and a macrostate with a single  $C$ , in volume  $V_{\sigma}$ . In fact, it can be shown that [55],

$$\begin{aligned} \Delta F_{\text{parity}} &= -k_B T \ln \left( \frac{K_{A+B}^{\text{eq}}}{V_{\sigma}} \right) \\ &= k_B T \ln \left( \frac{V_{\sigma}}{V_0} \right) + \Delta_{A+B} F^0, \end{aligned} \quad (29)$$

where  $\Delta_{A+B} F^0$  is the standard free energy of reaction, calculated in the reference volume  $V_0 = 1/\mathcal{C}_0$ . Thus the larger the system volume  $V_{\sigma}$ , the larger the standard free energy must be to give the same accuracy. If  $N_A^0$  is large,  $V_{\sigma}$  will also have to be large to ensure that the system is dilute and functions as intended. Therefore the standard free energy required for accurate computation will be very large and negative, and will become more negative logarithmically with system size at fixed initial concentration  $\mathcal{C}_A$ . The robustness of other molecular algorithms to finite reverse reaction rates, and possible ways to mitigate these effects, remain important open questions – although it seems likely that the least robust algorithms will be those that require a specific macrostate as an output.

## B. The meaning of thermodynamic reversibility

In Section VIA, I discussed microscopic reversibility – the concept that if a forwards reaction is possible, so is

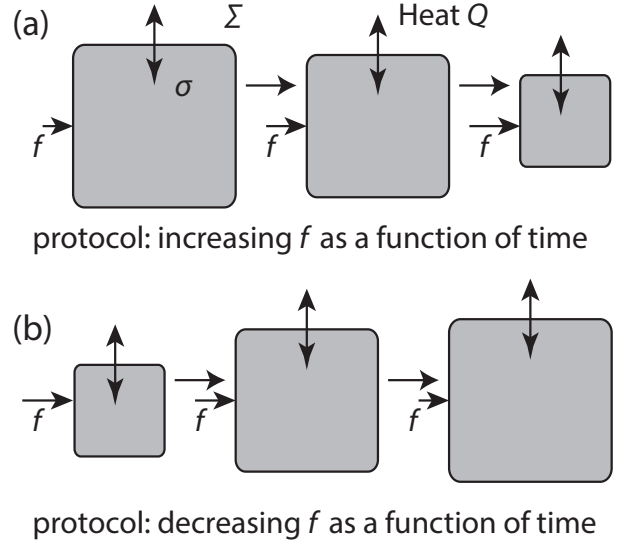


FIG. 10. The meaning of a reversible process. A schematic illustration of a process in which a time-dependent protocol (perhaps a compressing force) is applied to  $\sigma$ , leading to change in  $\sigma$  and exchange of heat with the environment  $\Sigma$ . The process is reversible if applying a time-reversed protocol as in (b) leads to both  $\sigma$  and  $\Sigma$  being returned to their initial conditions.

its microscopic reverse. This idea should *not* be conflated with the idea of thermodynamic reversibility, despite the unfortunate similarity of nomenclature. Thermodynamic reversibility is not the property of an individual transition or even a set of rate parameters describing a system. Rather, it is a property of an entire process in which an experimentalist or machine in the environment manipulates the system from the outside, applying some protocol (changing the conditions with time, as shown in Fig. 10(a)). If the system and environment are *both* be restored to their initial conditions by a time-reversed protocol (Fig. 10(b)), then the process and its time-reversed counterpart are thermodynamically reversible [97].

Reversible processes are important in thermodynamics because they do not increase the entropy of the universe, given by the sum of the entropy of  $\sigma$  and its environment  $\Sigma$  [97]. The entropy of the universe (an isolated system) cannot decrease, and excess entropy generation corresponds to wasted effort (or work). Hence reversible processes are the most efficient way of manipulating a system between a given start and end point. Historically, there has been significant interest in the minimal entropic cost of certain computational procedures, particularly the manipulation of a single bit, and the possibility of reversible computing [92, 98, 99].

It is often proposed that molecular systems might be candidates for the construction of reversible computers [64, 100, 101], and computational architectures with microscopically reversible dynamics have been analysed [100, 101]. Importantly, however, *microscopic reversibility does not imply thermodynamic reversibility*. In truth,



all systems possess microscopic reversibility – ignoring it is simply a modelling assumption, as discussed in Section VI A. Thermodynamic reversibility, however, depends on the initial conditions of the system, and the way in which a protocol is applied – not the intrinsic properties of a transition.

Consider, for example, a DNA-based toehold-exchange reaction (Fig. 6). We might be interested in switching the substrate strand  $A$  from a  $B$ -bound state to a  $C$ -bound state. If we set up the system to ensure that we start with a single  $AB$  duplex, and a separate  $C$  strand, and let it evolve naturally (i.e., implement a trivial protocol of “do nothing”) the system will undergo repeated strand exchange reactions, and the final state will alternate between  $AB$  and  $AC$ . This isn’t a full switch from  $AB$  to  $AC$ , but it is at least a change in the probability distribution over macrostates  $P_\sigma(i)$ .

During this process, both forwards and backwards reactions occur, and continue to occur indefinitely. So is this a thermodynamically reversible process? It is not. If we were to reverse our protocol of “doing nothing”, the system would not return to a state in which it was guaranteed to have an  $AB$  duplex; it would stay in an uncertain  $AB/AC$  state. Indeed, any overall change of a system (i.e., a change in the probability distribution  $P_\sigma(i)$ ) that occurs during a trivial “do nothing” protocol is *necessarily thermodynamically irreversible*, since the opposite change in  $P_\sigma(i)$  will not occur under a time-reversed “do nothing” protocol. The actual entropy increase in the process can be calculated using the methods discussed in Section VIII.

Note that it is not really meaningful to describe the individual reactions as reversible or irreversible in the thermodynamic sense. Once the system has relaxed to a completely uncertain equilibrium  $AB/AC$  state, strand exchange reactions will still occur from the perspective of individual trajectories. However,  $P_\sigma(i)$  will not change, and the subsequent (trivial) evolution of the system is reversible. It is only the initial period of transitioning from a guaranteed  $AB$  state to an uncertain  $AB/AC$  state that is irreversible – despite the fact that it involves the same microscopic strand exchange reactions. Fundamentally, thermodynamic reversibility is a property of the overall change of the system  $\sigma$  (through  $P_\sigma(i)$ ) and the environment  $\Sigma$ , rather than the individual molecular processes involved.

Does this observation prohibit reversible operations with molecular systems? In fact, it is possible to drive a change in the state of molecular systems reversibly. However, this must be done by externally changing the solution conditions in a quasistatic (slow) manner. For the above example, a reversible switch could be achieved by initially coupling the  $A$  strand to a buffer containing only  $B$  strands, and then replacing this buffer with a series of alternatives with gradually increasing ratios of  $C_C/C_B$ , as shown in Fig. 11. Eventually, in the limit  $C_C/C_B \rightarrow \infty$ , the switch to  $AC$  will be complete. By reversing the protocol, we return to the initial  $AB$  configuration, and

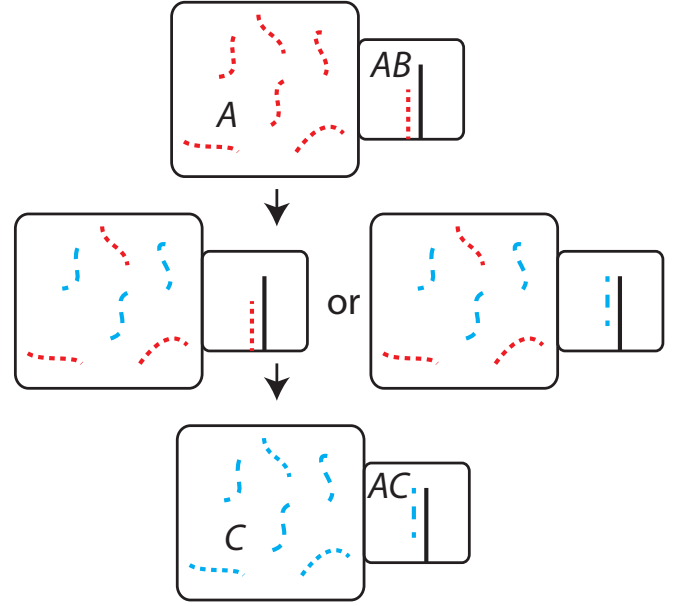


FIG. 11. A reversible switch of toehold-exchange system, achieved by coupling the volume containing strand  $A$  to buffers containing increasing amounts of strand  $C$ , and decreasing amounts of strand  $B$ . If this process is performed gradually enough, it is reversible.

restore the buffers to their original condition – hence the process is reversible. Similar procedures for reversible copying of a bit and reversible construction of a polymer copy from a template are discussed in more detail in Ref. [102] and Ref. [103], respectively. Although perhaps impractical, they highlight the difference between a thermodynamically reversible operation, and an irreversible process in which microscopic reversibility is relevant.

## VII. CATALYSIS AND THE CONSUMPTION OF MOLECULAR FUEL

The discussion in the previous sections has largely focused on systems that are isolated from the environment except for coupling to a heat bath. These systems are either in equilibrium, or would eventually reach equilibrium (although the required time might be extremely long). I also briefly discussed the idea of an experimenter manipulating system conditions, which is effectively changing the equilibrium state to which the system would relax, through a time-dependent protocol.

There is an important alternative setting, in which the system of interest is kept out of equilibrium by a continuous supply of chemical fuel. We might imagine that this fuel is supplied from an enormous buffer that is not depleted on the time scale of interest, or perhaps that the buffer is constantly topped up by the consumption of food (as occurs in cells). The prototypical example of a chemical fuel molecule is ATP, and so I will base my discussion around that example.



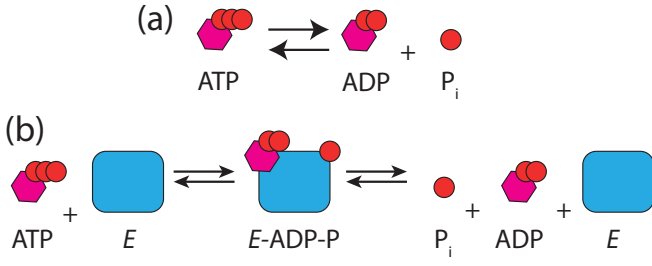
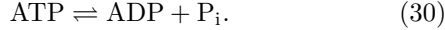


FIG. 12. (a) The breakdown of ATP into ADP and  $\text{P}_i$ . (b) a catalyst  $E$  can enhance the reaction rate by providing an alternative pathway; the catalyst is not consumed by the overall reaction.

ATP consists of a sugar-base group attached to a chain of three phosphate groups [62]. All of these bonds are covalent, but phosphate groups can be removed by hydrolysis. In particular, a key reaction is the removal of phosphate to give ADP (Fig. 12 (a)),



This reaction has an associated equilibrium constant  $K_{\text{ADP}+\text{P}_i}^{\text{eq}}$  and a standard free energy of ATP formation  $\Delta_{\text{ADP}+\text{P}_i} F^0$ . The cell uses the breakdown of food molecules to maintain a concentration imbalance of ATP, ADP and  $\text{P}_i$ ; *i.e.*,

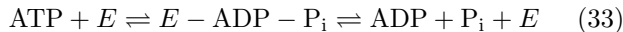
$$\frac{\mathcal{C}_{\text{ATP}}}{\mathcal{C}_{\text{ADP}}\mathcal{C}_{\text{P}_i}} > K_{\text{ADP}+\text{P}_i}^{\text{eq}}, \quad (31)$$

or equivalently

$$\begin{aligned} \Delta_{\text{ADP}+\text{P}_i} F_\sigma &= \mu_\sigma^{\text{ATP}} - (\mu_\sigma^{\text{ADP}} + \mu_\sigma^{\text{P}_i}) \\ &= \Delta_{\text{ADP}+\text{P}_i} F^0 + k_B T \ln \left( \frac{\mathcal{C}_0 \mathcal{C}_{\text{ATP}}}{\mathcal{C}_{\text{ADP}} \mathcal{C}_{\text{P}_i}} \right) \\ &> 0. \end{aligned} \quad (32)$$

Left in isolation, a solution of ATP, ADP and  $\text{P}_i$  prepared with these concentrations would relax to equilibrium by converting ATP into ADP and  $\text{P}_i$ . This process, which requires the disruption of covalent bonds, has extremely slow kinetics. Consequently the cell is able to build up and maintain a large concentration imbalance.

The kinetics of process such as that in Eq. 30 can be accelerated by catalysts [22, 62]. Catalysts provide an alternative reaction pathway between the endpoints, perhaps lowering the barriers associated with the formation and disruption of covalent bonds. A schematic illustration of catalyst operation is given in Fig. 12 (b), and one might record the effective reaction as



with  $E$  being the catalyst molecule. Importantly, the catalyst is released at the end of the process unchanged, and does not contribute to the free energy of the overall

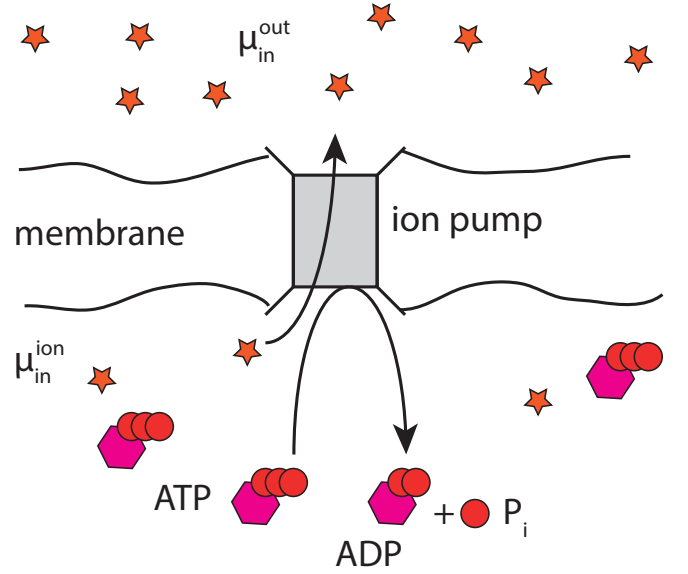


FIG. 13. An illustration of the principle of an ion pump. The pump systematically drives ions from inside the membrane to outside. By coupling the ion transition to  $\text{ATP} \rightleftharpoons \text{ADP} + \text{P}_i$ , ions can be systematically pumped against a chemical potential bias  $\mu_{\text{out}}^{\text{ion}} - \mu_{\text{in}}^{\text{ion}} \geq 0$ . This pumping is possible due to the large and negative contribution of the breakdown of ATP to the free energy difference associated with the transition.

reaction (its stoichiometric coefficient is zero). Therefore the catalyst does not affect the equilibrium balance between its substrate molecules, it only accelerates the rate at which this equilibrium is reached [22]. An equivalent way to view the action of a catalyst is that it accelerates *both* forwards and backwards rates equally.

In nature, some catalysts are simply present to accelerate reaction kinetics. For example, amylase is present to accelerate the breakdown of starch into sugars. In principle, it can also accelerate the conversion of sugars into starch, but this is irrelevant in the high starch conditions in which it operates. Many catalysts act on the ATP hydrolysis reaction in Eq. 30, but simply accelerating the equilibration of this process serves no function. Doing so would merely reduce the concentration imbalance built up by the cell. The purpose of these catalysts is to couple ATP hydrolysis to other reactions through the details of their internal biochemistry, and thereby use the free energy of ATP hydrolysis to drive those reactions in one way or another.

A clear example is given by ion pumps, molecular machines that reside in membranes, and transport ions across these membranes (Fig. 13). Without coupling to ATP hydrolysis, these ion pumps could only facilitate the equilibration of ions on either side of the membrane (*i.e.*, allow current to flow until the chemical potential of the ions is equal on both sides;  $\mu_{\text{in}}^{\text{ion}} = \mu_{\text{out}}^{\text{ion}}$ ). If, however, outwards ion transport is tightly coupled to ATP breakdown, then the overall free energy of pumping an

ion outward is

$$\Delta_{\text{in}} + \text{ATP} F_{\sigma} = \mu_{\text{out}}^{\text{ion}} - \mu_{\text{in}}^{\text{ion}} - (\mu_{\sigma}^{\text{ATP}} - (\mu_{\sigma}^{\text{ADP}} + \mu_{\sigma}^{\text{P}_i})), \quad (34)$$

which can be negative even if  $\mu_{\text{in}}^{\text{ion}} - \mu_{\text{out}}^{\text{ion}} > 0$ , due to the imbalance of chemical fuel molecules maintained by the cell. If the non-equilibrium chemical potentials of the ATP, ADP and  $\text{P}_i$  are maintained indefinitely, the ions will eventually reach a non-equilibrium steady state in which the outwards pumping is balanced by leaks back through other channels.

The non-equilibrium initial state of the fuel molecules is effectively a store of useful work. This store can be consumed to drive another process – in this case, the transfer ions against their chemical potential bias. The maximum chemical potential bias  $\mu_{\text{in}}^{\text{ion}} - \mu_{\text{out}}^{\text{ion}}$  against which progress can be made is simply the excess free energy change associated with breakdown of a single ATP,  $\mu_{\sigma}^{\text{ATP}} - (\mu_{\sigma}^{\text{ADP}} + \mu_{\sigma}^{\text{P}_i})$ .

Transferring ions across membranes is very much like charging a capacitor, and cells use these capacitors to drive other processes, including the firing of nerve cells. There are numerous other natural processes in which molecular fuel consumption is used to drive a coupled reaction. Molecular motors such as myosin, kinesin and dynein catalyse ATP breakdown to bias the direction in which they walk along a track [21, 22, 62]. Without coupling the forwards step to ATP hydrolysis, forwards and backwards steps would be equally likely, since all walker binding sites on the track have equal free energy by symmetry.

By coupling continuous fuel consumption to other molecular reactions via catalysts, it is therefore possible to drive those other reactions away from their natural equilibrium [104, 105]. A particularly important example are push-pull motifs, ubiquitous in the signalling mechanisms which pass information around the cell and illustrated in Fig. 14. In these small networks, the presence of active upstream catalysts leads to activation of downstream substrates, which can in turn propagate or respond to the signal.

Each push-pull motif consists of a protein that can be switched between its active ( $X^*$ ) and inactive ( $X$ ) states by binding of  $\text{P}_i$  to one (or more) amino acid residues. If this “phosphorylation” could only occur through binding and unbinding from  $\text{P}_i$  in solution, then the activity level would swiftly tend towards that determined by the equilibrium constant of binding and the abundance of  $\text{P}_i$ ,

$$\frac{C_{X^*}^{\text{eq}}}{C_X^{\text{eq}}} = C_{\text{P}_i} K_{X+\text{P}_i}^{\text{eq}} \quad (35)$$

Importantly, any upstream catalyst could only enhance this convergence to equilibrium, it could not change it. As discussed in Section IV A, relative equilibrium abundances are determined exclusively by the free energies of reactants and products – the concentration of a transiently involved catalysts is irrelevant.

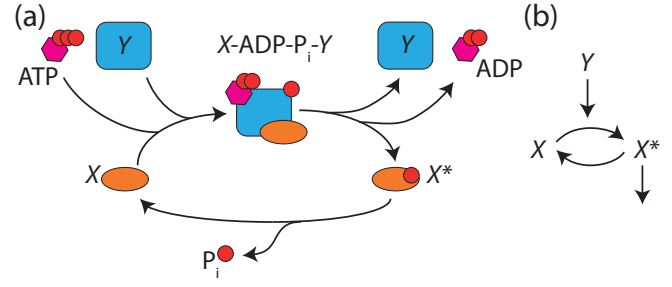
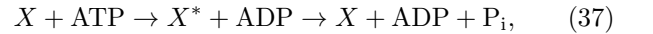


FIG. 14. A push-pull motif. The downstream molecule can  $X$  be converted between its phosphorylation states by two pathways; by exchanging phosphate with ADP/ATP or solution. Catalysts can accelerate one or both of these reactions; in (a), a catalyst  $Y$  is shown accelerating interactions with ATP/ADP. The overall negative free energy change of ATP breakdown favours the direction of reactions shown, rather than their reversed counterparts, and so  $X/X^*$  effectively undergoes a cycle. As a result, a signal related to the concentration of  $Y$  can be passed onto  $X^*$ , which then propagates the signal further. This function of the push-pull motif is illustrated in the simplified diagram in (b).

However, if ATP is present, phosphorylation can also occur via transfer of phosphate from ATP:



If the ATP, ADP and  $\text{P}_i$  concentrations are maintained such that  $\Delta_{\text{ADP}+\text{P}_i} F_{\sigma} > 0$ , it is impossible for the  $X/X^*$  subsystem to reach equilibrium. The series of reactions shown in Fig. 14 (a)



will necessarily occur much more frequently on sample trajectories than the counterpart



The reactions in Eq. 37 will occur more frequently since the net result of the reactions in Eq. 37 is the breakdown of ATP, whereas the net result of the reaction in Eq. 38 is the synthesis of ATP from ADP and  $\text{P}_i$ , and the sign of  $\Delta_{\text{ADP}+\text{P}_i} F_{\sigma} = \mu_{\sigma}^{\text{ATP}} - (\mu_{\sigma}^{\text{ADP}} + \mu_{\sigma}^{\text{P}_i}) > 0$  favours ATP breakdown. The protein  $X$  systematically tends to be converted into  $X^*$  by one pathway, and converted back into  $X$  by a completely different pathway (Fig. 14 (a)). From the perspective of  $X/X^*$ , detailed balance appears to be violated, and therefore the coupling to chemical fuel drives the  $X/X^*$  subsystem out of equilibrium. As a result, when the hydrolysis-driven cycle of Eq. 37 becomes dominant relative to Eq. 38, the steady state ratio  $C_{X^*}/C_X$  is determined purely by the relative rates of the two reactions within that cycle,  $X + \text{ATP} \rightarrow X^* + \text{ADP}$  and  $X^* + \text{ADP} \rightarrow X + \text{ADP} + \text{P}_i$ . This is analogous to the way in which the behaviour of cellular mRNA can be analysed whilst ignoring the microscopic reverse process to transcription (Section VI A).

Since the dominant reactions that interconvert  $X$  and  $X^*$  are not a microscopic reverse pair, thermodynamics places no restrictions on their relative rates. In particular, it is perfectly possible for a catalyst to accelerate the kinetics of one reaction and not the other. The output ratio  $C_{X^*}/C_X$  can then be sensitive to the concentration of an upstream catalyst ( $Y$  in Fig. 14(a)) that accelerates the exchange of phosphate between  $X$  and ATP. In this manner, signalling cascades can pass on information on the concentration of active upstream catalysts, illustrated schematically in Fig. 14(b), as required. I emphasize that it would be impossible for  $Y$  to influence the  $X/X^*$  ratio in this way without consuming fuel, or binding to the downstream molecule and remaining bound (*ie.*, with  $Y$  itself being consumed by the reaction).

The above catalytic activation motif is extremely powerful. Since the upstream molecule acts catalytically, it is not consumed in the act of passing on the signal. It is thus able to interact with other downstream proteins to either amplify or branch the signal [65]. Moreover, the downstream readout's persistent modification allows it to keep a long-lived record of the state of the upstream protein, permitting time-integration of signals [66, 67]. These features are only possible in a motif that incorporates catalytic signal propagation, which in turn relies on coupling to a non-equilibrium fuel source.

Clearly, driving of reactions by coupling them to a non-equilibrium fuel supply is a central motif in natural systems [22, 104, 105]; other examples include enhancement of substrate selectivity through “kinetic proofreading” [106, 107], important in replication, transcription and translation of nucleic acids; and maintenance of oscillations related to cell cycle [108] and circadian clocks [63]. Needless to say, these fuel-consuming processes are widely exploited in synthetic biology, and enzyme-driven processes in cell-free environments [27, 28, 68]. There has also been some exploration of these ideas in nucleic acid nanotechnology, most notably in the design of autonomous DNA walkers [109–115], some of which are powered by base pairing alone [115–117]. In particular, hairpins can be used as metastable non-equilibrium fuel [80, 116, 117]; the motion of the walker is coupled to catalysing reactions that are otherwise frustrated by the hairpin structure.

More generally, recent work has shown how to implement arbitrary chemical reaction networks (CRNs) as nucleic acid systems by realising each reaction as a multi-stage process involving ancillary molecules [74, 96]. These ancillary molecules can function as fuel, in principle allowing catalytic driving out of equilibrium as discussed above [74]. Most impressively, the Khammash group using a nucleic-acid-based architecture to implement a CRN-based noise filter powered by fuel consumption [35]. When considering the capabilities and design possibilities of CRNs, it is always worth noting whether non-equilibrium fuel is required to provide a large thermodynamic bias for the system to function as intended. Furthermore, when these implicit fuel molecules are sys-

tematically consumed by the system, detailed balance will in general be violated for the other species considered explicitly.

Notwithstanding the above examples, the common natural motif of a network powered by fuel consumption is relatively rare in non-enzymatic artificial nanotechnology. Given the uses of fuel-powered systems, producing artificial analogues is an obvious goal. One major advantage would be that, given a constant supply (or sufficiently large buffer) of fuel, such systems could potentially operate indefinitely, performing repeated operations, rather than functioning as single-shot devices. Continuous operation would be essential for implementing circuits that perform functions like feedback-control [95].

## VIII. STOCHASTIC THERMODYNAMICS AND ITS RELEVANCE TO BIOMOLECULAR SYSTEMS

### A. Generalising thermodynamic quantities to non-equilibrium distributions

Thermodynamic quantities, as introduced in Section II A, describe average properties of systems in equilibrium. In particular, the internal energy, entropy and free energy in Eqs. 3, 4 and 5 are defined with respect to the Boltzmann distribution. I then introduced the idea of free energies, energies and entropies of macrostates,  $F_\sigma(i)$ ,  $U_\sigma(i)$  and  $S_\sigma(i)$ , defined analogously but with the integral over degrees of freedom restricted to that macrostate. These quantities, particularly  $F_\sigma(i)$ , were very helpful in analysing the typical states in equilibrium, in constraining the dynamics of the underlying stochastic process, and in understanding how molecular fuel molecules can power reactions. However, they do not directly correspond to non-equilibrium versions of  $F_\sigma^{\text{eq}}$ ,  $U_\sigma^{\text{eq}}$  and  $S_\sigma^{\text{eq}}$ . Given a non-equilibrium distribution over macrostates, I have provided no definition of the free energy, average internal energy or entropy of the system.

This absence is important, because in molecular contexts we are typically interested in systems which evolve from one non-equilibrium distribution to another, or relax from a non-equilibrium distribution to equilibrium. To actually discuss the thermodynamics of these processes, rather than simply using  $F_\sigma(i)$  to constrain the kinetics, it is necessary to generalise  $F_\sigma^{\text{eq}}$ ,  $U_\sigma^{\text{eq}}$  and  $S_\sigma^{\text{eq}}$  to systems that exhibit large deviations from equilibrium. Such a generalisation [118, 119] will permit the evaluation of thermodynamics of non-equilibrium processes, allowing estimation of quantities such as efficiency, for example.

The obvious generalisation of the internal energy to an arbitrary distribution  $P_\sigma(\mathbf{x}, \mathbf{p})$  is

$$\mathcal{U}_\sigma[P_\sigma(\mathbf{x}, \mathbf{p})] = \int_{\mathbf{x}, \mathbf{p}} d\mathbf{x}, d\mathbf{p} E_\sigma(\mathbf{x}, \mathbf{p}) P_\sigma(\mathbf{x}, \mathbf{p}), \quad (39)$$

the average of the microstate energy over  $P_\sigma(\mathbf{x}, \mathbf{p})$ . If

the Markovian approximation for macrostate dynamics holds, then by definition the system is well-equilibrated within macrostates, even if different macrostates have non-equilibrium probabilities  $P_\sigma(i) \neq P_\sigma^{\text{eq}}(i)$ . In this case,

$$\mathcal{U}_\sigma[P_\sigma(i)] = \sum_i P_\sigma(i) U_\sigma(i). \quad (40)$$

We also generalise the entropy in the same way, since the uncertainty in a non-equilibrium distribution remains well-defined

$$\mathcal{S}_\sigma[P_\sigma(\mathbf{x}, \mathbf{p})] = -k_B \int_{\mathbf{x}, \mathbf{p}} d\mathbf{x}, d\mathbf{p} P_\sigma(\mathbf{x}, \mathbf{p}) \ln(P_\sigma(\mathbf{x}, \mathbf{p})/\rho). \quad (41)$$

When considering well-defined macrostates, this expression becomes

$$\mathcal{S}_\sigma[P_\sigma(i)] = -k_B \sum_i P_\sigma(i) \ln P_\sigma(i) + \sum_i P_\sigma(i) S_\sigma(i). \quad (42)$$

The generalisation is slightly more complex than for the internal energy; we obtain a term corresponding to the average entropy of the macrostates *and* a term corresponding to the uncertainty in the macrostate. Similarly, we can generalise the free energy:

$$\begin{aligned} \mathcal{F}_\sigma[P_\sigma(i)] &= \mathcal{U}_\sigma[P_\sigma(i)] - T\mathcal{S}_\sigma[P_\sigma(i)] \\ &= \sum_i P_\sigma(i) F_\sigma(i) + k_B T \sum_i P_\sigma(i) \ln P_\sigma(i). \end{aligned} \quad (43)$$

With these definitions, the total entropy production of the system and environment during a process in which  $P_\sigma(i)$  evolves to  $P'_\sigma(i)$  is

$$\begin{aligned} T\Delta\mathcal{S}_{\sigma+\Sigma} &= T\Delta\mathcal{S}_\Sigma + T\mathcal{S}_\sigma[P'_\sigma(i)] - T\mathcal{S}_\sigma[P_\sigma(i)] \\ &= -(\mathcal{U}_\sigma[P'_\sigma(i)] - \mathcal{U}_\sigma[P_\sigma(i)]) \\ &\quad + T\mathcal{S}_\sigma[P'_\sigma(i)] - T\mathcal{S}_\sigma[P_\sigma(i)] \\ &= -(\mathcal{F}_\sigma[P'_\sigma(i)] - \mathcal{F}_\sigma[P_\sigma(i)]). \end{aligned} \quad (44)$$

In the above, I have used the fact that a simple chemical system can only exchange energy in the form of heat with its environment. Thus the entropy increase of the environment due to the process is given directly by the reduction in internal energy of  $\sigma$  [9, 10, 71]:  $\Delta\mathcal{S}_\Sigma = -(\mathcal{U}[P'(i)] - \mathcal{U}[P(i)])/T$ . In more complex environments, alternative results are obtained which also included the external work done on the system [9, 10, 71].

Remarkably, as I outline in Section VIII B, the total generalised entropy of Eq. 44 is guaranteed to be a non-decreasing function of time provided detailed balance always holds. We can then state a second law for non-equilibrium molecular processes:

$$T\Delta\mathcal{S}_{\sigma+\Sigma} = -(\mathcal{F}_\sigma[P'_\sigma(i)] - \mathcal{F}_\sigma[P_\sigma(i)]) \geq 0. \quad (45)$$

Eq. 45 is highly significant. Firstly, it emphasizes the importance of the (generalised) free energy as a thermodynamic resource in simple molecular systems. Since

total  $\mathcal{F}_\sigma[P_\sigma(i)]$  can only decrease, if a process acts to increase  $\mathcal{F}_\sigma[P_\sigma(i)]$  for a subset of species, a compensatory decrease in  $\mathcal{F}_\sigma[P_\sigma(i)]$  must occur elsewhere. This observation generalises the discussion in Section VII on the use of chemical fuel molecules to drive other components of a system out of equilibrium.

Generalised thermodynamic quantities also allow us to meaningfully analyse the entropy generation (and hence irreversibility) of processes such as that considered in Section VI A, when a duplex initially prepared in the  $AB$  state relaxes to an equilibrium of equal probability  $AB/AC$ , via repeated rounds of strand exchange. If the equilibrium distribution has  $AB$  and  $AC$  duplexes with equal probability (I assume that the strands are dilute enough that three-stranded complex is rarely observed), then both macrostates have equal free energy. Thus  $\sum_i P_\sigma(i) F_\sigma(i)$  is equal in the initial and final states. Any change in generalised free energy can only arise from the second term in Eq. 43, the difference in generalised entropy due to the distribution over macrostates. Indeed, for a system initially guaranteed to be in macrostate  $AB$ ,  $\sum_i P_\sigma(i) \ln P_\sigma(i) = 0$ ; whereas the uncertainty in the final macrostate gives  $-\sum_i P'_\sigma(i) \ln P'_\sigma(i) = \ln 2$ . Consequently

$$T\Delta\mathcal{S}_{\sigma+\Sigma} = -(\mathcal{F}_\sigma[P'_\sigma(i)] - \mathcal{F}_\sigma[P_\sigma(i)]) = k_B T \ln 2 > 0 \quad (46)$$

for the irreversible strand exchange process discussed in Section VI A. Note that the full definition of the generalised free energy, including the term arising from the entropy of the distribution over macrostates, is necessary to obtain this result.

Perhaps even more interesting is the case when neither the initial *nor* the final distribution correspond to equilibrium. For example, a self-assembling structure need not reach the equilibrium distribution after a finite time, implying  $\mathcal{F}_\sigma[P'_\sigma(i)] > \mathcal{F}_\sigma[P_\sigma^{\text{eq}}(i)]$ , since the equilibrium state minimises the generalised free energy by definition. But since  $\mathcal{F}_\sigma[P_\sigma(i)]$  can only decrease with time (Eq. 45), producing a non-equilibrium distribution necessarily requires a higher initial generalised free energy than producing an equilibrium distribution. In other words, producing the non-equilibrium distribution has a greater minimal resource cost.

Recent work has analysed how higher initial free energies allow self-assembly of non-equilibrium structures [120, 121]. Intriguingly, these non-equilibrium assemblies can have a lower density of structural defects than in equilibrium [120], suggesting a possible alternative to optimising assembly by encouraging the approach to equilibrium. Indeed, it has recently been shown that a related process, the production of polymer copies that persist after separation from their templates (as occurs in replication, transcription and translation in cells) is inherently an exercise in producing out-of-equilibrium structures. In this context, an equilibrium output is unrelated to its template, and so a high free energy initial state is required to give any accuracy at all [103].

## B. Individual trajectories and the second law

Arguably the deepest result of stochastic thermodynamics is “deriving” the second law (Eq. 45) from an assumption of detailed balance. Moreover, the second law inequality follows from a *fluctuation relation* equality for a quantity that can be interpreted as a trajectory-dependent entropy, highlighting the statistical nature of entropy. I will now briefly outline this derivation for the relevant case of a simple chemical system that can only exchange heat with its environment. The general case is more complicated, but the ideas are similar [9, 10, 71]. To do so, I will introduce quantities that depend on the specific stochastic trajectory taken by the system – these quantities give stochastic thermodynamics its name.

Let  $z(t)$  be a sample trajectory generated by the underlying Markovian dynamics in the space of macrostates of the system. Let its initial value be  $z(0)$  and its final value be  $z(\tau)$ , then define  $P_\sigma([z(t)]|z(0))$  as the probability of observing that trajectory given the initial condition, and  $P_\sigma([\tilde{z}(t)]|z(\tau))$  as the probability of observing exactly the time-reversed trajectory given a starting point of  $z(\tau)$ . The fundamental assumption underlying stochastic thermodynamics, for our simple molecular system coupled only to a heat bath, is

$$k_B \ln \frac{P_\sigma([z(t)]|z(0))}{P_\sigma([\tilde{z}(t)]|z(\tau))} = \Delta S_\Sigma[z(t)] + (S_\sigma(z(\tau)) - S_\sigma(z(0))) = -(F_\sigma(z(\tau)) - F_\sigma(z(0)))/T. \quad (47)$$

Here,  $\Delta S_\Sigma[z(t)] = -(U(z(\tau)) - U(z(0)))/T$  is the entropy increase in the environment due to heat transfer from  $\sigma$  along this particular trajectory.  $F_\sigma(z)$ ,  $S_\sigma(z)$  and  $U_\sigma(z)$  are the macrostate free energy, entropy and energy defined in Section VIII A (*ie.*, not the generalised quantities). The assumption in Eq. 47 is automatically satisfied if detailed balance holds (Eq. 22), since if  $k_{ji}/k_{ij} = \exp(-(F_\sigma(j) - F_\sigma(i))/k_B T)$ , then the overall probability of a trajectory and its time-reverse will also be given by the total free energy difference. To see this fact, note that a time-reversed trajectory occupies the same states for the same length of time, but simply makes the reverse transitions; the relative probability of the two trajectories is then the product of the relative rates of the reverse transitions.

To proceed, let us define the “trajectory-dependent” entropy production [9, 10, 71]:

$$\Delta s[z(t)] = \Delta S_\Sigma[z(t)] + (S_\sigma(z(\tau)) - S_\sigma(z(0))) - k_B \ln \frac{P_\sigma(z(\tau))}{P_\sigma(z(0))}. \quad (48)$$

This exotic quantity, when averaged over all possible trajectories  $z(t)$ , will give the generalised entropy change of the entire process – hence its name. To see that  $\sum_{z(t)} P_\sigma[z(t)] \Delta s[z(t)] = \Delta S_{\sigma+\Sigma}$ , compare Eq. 48 to

Eq. 42 and 44. Note that Eq. 48 contains a term for the entropy of the environment  $\Sigma$ , a term for the change in macrostate entropy  $S_\sigma$ , and a term related to the distribution of  $\sigma$  over its macrostates, as required.

Combining Eq. 47 and Eq. 48, we see that

$$\frac{P_\sigma[z(t)]}{P_\sigma[\tilde{z}(t)]} \exp(-\Delta s[z(t)]/k_B) = 1. \quad (49)$$

Here,  $P_\sigma[z(t)]$  is the probability of observing the trajectory  $z(t)$ , including the initial probability of being at  $z(0)$ , and  $P_\sigma[\tilde{z}(t)]$  is the probability of observing the reverse trajectory, including a distribution of initial states of the reverse trajectories given by  $P_\sigma(z(\tau))$ . We can multiply by  $P_\sigma[\tilde{z}(t)]$  and sum over all possible trajectories, yielding

$$\sum_{z(t)} P_\sigma[z(t)] \exp(-\Delta s[z(t)]/k_B) = \sum_{\tilde{z}(t)} P_\sigma[\tilde{z}(t)] = 1, \quad (50)$$

in which I have used the fact that a sum over all trajectories  $z(t)$  is equivalent to a sum over all reverse trajectories  $\tilde{z}(t)$ . We have arrived at the celebrated fluctuation relation for entropy [9, 10, 71],

$$\langle \exp(-\Delta s[z(t)]/k_B) \rangle = 1, \quad (51)$$

where the average is defined over all possible trajectories  $z(t)$ . The conventional second law then follows using Jensen’s inequality  $\ln \langle \exp(f(v)) \rangle \leq \langle f(v) \rangle$ , which implies

$$\begin{aligned} \langle \Delta s[z(t)]/k_B \rangle &= \Delta S_\Sigma[z(t)] + S_\sigma[P_\sigma(\tau)] - S_\sigma[P(0)] \\ &= \Delta S_{\sigma+\Sigma}[z(t)] \\ &= -(\mathcal{F}_\sigma[P'_\sigma(i)] - \mathcal{F}_\sigma[P_\sigma(i)]) \geq 0, \end{aligned} \quad (52)$$

or equivalently that the total entropy of system and environment cannot decrease.

The fluctuation relation is a remarkable result, implying that at the level of individual trajectories, “negative entropy” paths are possible but statistically unfavoured. The result effectively shows that the familiar second law follows from a more general statement about the statistical properties of trajectories, which in turn follows directly from the relatively simple assumption of detailed balance. As a result, it emphasises the statistical interpretation of entropy and the second law, and their natural emergence from familiar stochastic dynamics. A host of alternative fluctuations can also be derived, depending on the context [9, 10, 71].

Many of the results of stochastic thermodynamics, including the fluctuation relation itself, are arguably more philosophically deep than immediately useful. However, the fluctuating, far-from-equilibrium nature of many biochemical systems – particularly at the single molecule level – often lends itself to analysis using these tools. Generalisations including the possibility of applying external work have been used to study molecular systems undergoing non-equilibrium stress-induced transitions in

experiment [122, 123]. On a more theoretical level, an extremely common approach is to use imbalances in probability flows to infer the entropy generation (or free-energy consumption) of a molecular network [64, 124–126]. The entropy cost is paid by the consumption of a (typically implicit) molecular fuel molecule. Thus the resource cost of various molecular networks, including sensing, adaptive information-processing and force-generating systems, can be estimated and any trade-offs between performance and cost explored.

One important feature of stochastic thermodynamics is the ability to relate the relative probabilities of entire trajectories and their microscopic reverses to entropy generation (Eq. 47). This allows the simultaneous analysis and comparison of the entropy cost of distinct trajectories that move between macrostates via different pathways of different lengths. England has used this formalism to argue for a minimal bound on the entropy generation of replicators related to the overall birth and death rates [127]. Similarly, by considering the entropy generated by simple and more complex activation pathways, it was recently demonstrated that single-step activation is optimally efficient for signal-propagating push-pull networks of the type shown in Fig. 14(b) [102].

## IX. CONCLUSIONS

In this pedagogical perspective, I have attempted to outline the manner in which molecular thermodynamics impacts the behaviour of both natural and artificial biochemical systems. Starting from the basic equilibrium statistical mechanics of dilute molecular systems, I introduced the concept of biochemical macrostates and the free energy of a macrostate. I then discussed the stochastic kinetics of molecular systems, and the constraints on kinetics implied by the macrostate free energies and detailed balance - which are strong, but insufficient to fully specify dynamics. Finally, I discussed how stochastic thermodynamics allows for a consistent thermodynamic interpretation of the stochastic evolution of a far-from-equilibrium system, both at the level of individual trajectories and the probability distribution as a whole.

Simultaneously, I have tried to highlight the relevance of these concepts to molecular systems that are typically studied and engineered. Thermodynamic ideas dictate both fundamental principles and general limits of what is possible, as well as contributing to detailed mechanistic understanding and design. In particular, defining biochemical macrostates allows for modelling of biochemical equilibria, and the state of a self-assembling system that would be expected in the limit of infinitely long time. In finite time, however, a system may not even come close to equilibrium - it is therefore important to understand kinetics. Kinetics are also relevant in systems that do not reach equilibrium because they are coupled to a supply of molecular fuel. The operation of such systems, which

are widespread in nature but currently less common in artificial contexts, can be understood from the overriding tendency of the fuel molecules to relax towards their equilibrium. Finally, by defining entropies and free energies of trajectories and non-equilibrium distributions, stochastic thermodynamics allows for an analysis of the resource cost of various far-from-equilibrium processes that are important to both natural and artificial systems.

At the same time, I have emphasised underlying assumptions that are often given insufficient consideration, and common misconceptions. For example, species concentration is an important component in the stability of self-assembling structures; a “melting temperature” should never be defined without reference to concentration, and standard free energies are not directly meaningful at reasonable strand concentrations. Markov models at the level of biochemical macrostates are only appropriate if the macrostates are well-defined so that transitions between them are rare events. Kinetic models of systems that are not subject to driving should exhibit detailed balance in equilibrium. This fact is sometimes ignored without careful consideration of the consequences; the assumption of perfect irreversibility can have major effects, depending on the context. Finally, catalysts can only influence the yield of downstream substrates if catalysis is coupled to a supply of non-equilibrium fuel.

The flow of understanding isn’t solely in one direction, however. Biomolecular systems can also contribute to the understanding of fundamental thermodynamics. Most obviously, although stochastic thermodynamics wasn’t developed solely for molecular systems, it finds perhaps its most natural application there. Important processes in both natural and artificial molecular systems involve single molecules undergoing large fluctuations that are essential to function. Living systems are, almost by definition, kept in a far from equilibrium state. Pursuing important questions in such systems will thus drive understanding of the underlying thermodynamic principles, and the techniques used to study them. For example, many authors are currently probing the deep connections between information theory and (stochastic) thermodynamics in a quest to understand the limits and capabilities of cellular sensing, signalling and adaption networks [67, 102, 128, 129].

Much of the difficulty in understanding fundamental thermodynamics stems from interpreting the results, rather than from mathematical complexity. This is partly because many of the concepts, beginning with entropy itself, are inherently abstract. As a result, debate persists about the proper interpretation of relatively fundamental systems and concepts [130–133] – debates that I have ignored in this perspective. Biomolecular systems, however, are inherently concrete – even when modelled in a simplistic manner. They can thus serve to demystify the debate. As an example, there has been much recent interest in the possibility of designing an engine that exploits 1s and 0s on a tape to do useful work (such as lifting a weight) [134, 135]. When the state of the tape is writ-



ten as abstract 1s and 0s, the operation of such a device seems almost magical. However, for the device to run, the 1s and 0s must interact with a motor in a specific way; they must be physical. And a physical instantiation of the system can be constructed from biomolecules – at which point it becomes clear that the 1s and 0s correspond to two states of a molecular fuel molecule, and the motor is driven by an imbalance of fuel input *just like any other motor* [136]. The operation of the device is consistent with well-established laws of thermodynamics, rather than being particularly remarkable.

By constraining ourselves to physically plausible models of molecular systems, it is also much clearer what is possible. It is consequently harder to accidentally invent an unaccounted-for “Maxwell Demon” that performs an impossible task, or miss the requirements of a key step. Moreover, there is at least some sense in which the practical constraints of thought experiments become evident. A thermodynamic system can be designed on paper with arbitrary coupling between degrees of freedom, but when trying to instantiate it as a molecular system, even theoretically, the inherent trade-offs required become more obvious [136].

In particular, the majority of thermodynamic studies focus on systems in which changes are driven by intervention from an external experimenter, who does work on the system (this idea is briefly introduced in Section II and VIB). From the perspective of stochastic thermodynamics, doing work corresponds to adjusting the system to change the energy of the microstates in a time-dependent manner. Whilst analysis in terms of external work is not wrong, it is again hard to understand what is fundamentally possible at zero cost, particularly for small systems. Generally, the actual physical mechanism by which work is applied is not considered; it is therefore unclear which work protocols are feasible, and whether work can actu-

ally be applied, recovered and stored efficiently [137], as is assumed. For example, recent experiments on the manipulation of single colloids actually implement “work” protocols through a highly dissipative mechanism, the cost of which far exceeds any entropy generated by the motion of the colloid itself (which is typically the topic of interest) [138, 139]. Similarly, in questions related to the minimal cost of operations on small systems, such as computing with a single bit, the cost of an experimenter deciding to implement each stage of the protocol is not obviously accounted for [137].

By contrast, molecular systems are usually thought of as running autonomously, without external work (although this is not always the case [102, 103]). Due to the ability of molecules to diffuse and interact selectively with multiple partners, complex behaviour can arise simply from a system initiated in some non-equilibrium state and left to evolve (perhaps with a constant supply of non-equilibrium fuel). In these autonomous systems, all costs are explicit, and all behaviours based on plausible chemical reactions. A major challenge for the future of thermodynamics lies in understanding the fundamental differences between the capabilities of such autonomous systems, and those systems in which outside manipulation is permitted. In turn, understanding these autonomous systems will potentially allow us to build microscopic devices of extremely high efficiency that genuinely approach the fundamental lower bounds on cost [102].<sup>1</sup>

## ACKNOWLEDGMENTS

The author wishes to acknowledge fruitful conversations with M. Gopalkrishnan, P. R. ten Wolde, A. A. Louis, J. P. K. Doye, M. Gopalkrishnan and A. J. Turberfield. This work was supported by a Royal Society University Research Fellowship.

- 
- [1] N.L.S. Carnot, E. Clapeyron, R. Clausius, *Reflections on the motive power of fire - and other papers on the second law of thermodynamics* (Dover, New York, 1960)
  - [2] J.W. Gibbs, *The Scientific Papers of J. Willard Gibbs - Volume One* (Longmans, Green and Co., London, 1906)
  - [3] L. Boltzmann, *Lectures on gas theory* (Dover, New York, 1964)
  - [4] J.M. Yeomans, *Statistical mechanics of phase transitions* (Oxford University Press, Oxford, 1992)
  - [5] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambrid, Cambridge, UK, 1996)
  - [6] D. Frenkel, B. Smit, *Understanding Molecular Simulation* (Academic Press Inc. London, 2001)
  - [7] M.E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, Oxford, 2010)
  - [8] G.E. Crooks, Phys. Rev. E **60**, 2721 (1999)
  - [9] C. Jarzynski, Annu. Rev. Cond. Matt. Phys. **2**, 329 (2011)
  - [10] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012)
  - [11] G.J. Mulder, Journal für praktische Chemie **16**, 129 (1839)
  - [12] M. Teich, D.M. Needham, *A documentary history of biochemistry, 1770-1940* (Faileigh Dickinson University Press, Rutherford, NJ, 1992)
  - [13] J.B. Sumner, J. Biol. Chem. **69**, 435 (1926)
  - [14] J. Kendrew, G. Bodo, H. Dintzia, R. Parrish, H. Wyckoff, D. Phillips, Nature **181**, 662 (1958)
  - [15] H. Muirhead, M. Perutz, Nature **199**, 633 (1963)
  - [16] O.T. Avery, C.M. MacCleod, M. McCarty, J. Exp. Med. **79**(2), 137 (1944)
  - [17] J.D. Watson, F.H.C. Crick, Nature **171**, 737 (1953)
  - [18] F.H.C. Crick, *Symposia of the Society for Experimental Biology, Number XII: The Biological Replication of Macromolecules* (Cambridge University Press, 1958), chap. On protein synthesis, pp. 138-163
  - [19] F.H.C. Crick, Nature **227**, 561 (1970)
  - [20] U. Alon, *Introduction to Systems Biology: Design Prin-*

- ciples of Biological Networks* (CRC press, Boca Raton, FL, 2007)
- [21] J. Howard, *Mechanics of motor proteins and the cytoskeleton* (Sinauer Associates, Inc., Sunderland, MA, 2001)
  - [22] P. Nelson, *Biological physics: Energy, information, life* (W. H. Freeman and Co., New York, NY, 2004)
  - [23] G. Baldwin, T. Bayer, R. Dickinson, T. Ellis, P.S. Freemont, R.I. Kitney, K. Polizzi, G.B. Stan, *Synthetic biology – a primer* (Imperial College Press, London, UK, 2012)
  - [24] Y.J. Chen, B. Groves, R. Muscat, G. Seelig, *Nat. Nanot* **10**, 748 (2015)
  - [25] Y. Hsia, J.B. Bale, S. Gonen, D. Shi, W. Sheffler, K.K. Fong, U. Nattermann, C. Xu, P.S. Huang, R. Ravichandran, S. Yi, T.N. Davis, T. Gonen, N.P. King, D. Baker, *Nature* pp. 136–139 (2016)
  - [26] J.B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, D. Thomas, C. and Cascio, T.O. Yeates, T. Gonen, N.P. King, D. Baker, *Science* **353**, 389 (2016)
  - [27] T. Fujii, Y. Rondelez, *ACS nano* **7**, 27 (2012)
  - [28] A.A. Green, P.A. Silver, J. Collins, P. Yin, *Cell* **159**, 925 (2014)
  - [29] P.W.K. Rothmund, *Nature* **440**, 297 (2006)
  - [30] S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W.M. Shih, *Nature* **459**, 414 (2009)
  - [31] Y. Ke, L.L. Ong, W.M. Shih, P. Yin, *Science* **338**, 1177 (2012)
  - [32] E. Winfree, F.R. Liu, L.A. Wenzler, N.C. Seeman, *Nature* **394**, 539 (1998)
  - [33] G. Seelig, D. Soloveichik, D.Y. Zhang, E. Winfree, *Science* **314**, 1585 (2006)
  - [34] L. Qian, E. Winfree, *Science* **332**, 1196 (2011)
  - [35] C. Zechner, G. Seelig, M. Rullan, M. Khammash, *Proc. Nat. Acad. Sci. USA* **113**, 4729 (2016)
  - [36] R. Kitney, P. Freemont, *FEBS lett.* **586** (2012)
  - [37] V. Chubukov, A. Mukhopadhyay, C.J. Petzold, J.D. Keasling, H.G. Martín, *NPJ Syst. Biol. Appl.* **2**, 16009 (2016)
  - [38] K. Huang, *Statistical Mechanics, Second Edition* (John Wiley & Sons, Inc., New York, 1987)
  - [39] M. Orozco, A. Pérez, A. Noy, F.J. Luque, *Chem. Soc. Rev.* **32**, 350 (2003)
  - [40] E.A. Cino, W.Y. Choy, M. Karttunen, *J. Chem. Theory Comput.* **8**, 2725 (2012)
  - [41] S. Marrink, H. Risselada, S. Yefimov, D. Tieleman, A. de Vries, *J. Phys. Chem. B* **111**, 7812 (2007)
  - [42] J.P.K. Doye, T.E. Ouldridge, A.A. Louis, F. Romano, P. Šulc, C. Matek, B. Snodin, L. Rovigatti, J.S. Schreck, R.M. Harrison, W.P.J. Smith, *Phys. Chem. Chem. Phys.* **15**, 20395 (2013)
  - [43] E.T. Jaynes, *Phys. Rev.* **106**, 620 (1957)
  - [44] C.E. Shannon, W. Weaver, *The mathematical theory of communication* (University of Illinois Press, 1949)
  - [45] M. Nakata, G. Zanchetta, B.D. Chapman, C.D. Jones, J.O. Cross, R. Pindak, T. Bellini, N.A. Clark, *Science* **318**(5854), 1276 (2007)
  - [46] C. De Michele, L. Rovigatti, T. Bellini, F. Sciortino, *Soft Matter* **8**, 8388 (2012)
  - [47] J. Chen, N.C. Seeman, *Nature* **350**, 631 (1991)
  - [48] J.P.J. Sobczak, T.G. Martin, T. Gerling, H. Dietz, *Science* **338**, 1458 (2012)
  - [49] S.E. Ahnert, J.A. Marsh, H. Hernandez, C.V. Robison, S.S. Teichmann, *Science* **350**, aaa2245 (2015)
  - [50] H. Fraenkel-Conrat, R.C. Williams, *Proc. Natl. Acad. Sci. USA* **41**(10), 690 (1955)
  - [51] I.G. Johnston, A.A. Louis, J.P.K. Doye, *J. Phys.: Condens. Matter* **22**(10), 104101 (2010)
  - [52] R.P. Goodman, I.A.T. Sharp, C.F. Tardin, C.M. Erben, R.M. Berry, C.F. Schmidt, A.J. Turberfield, *Science* **310**, 1661 (2005)
  - [53] G. Tikhomirov, P. Petersen, L. Qian, *Nat. Nanotechnol.* (2017)
  - [54] A. Murugan, J. Zou, M.P. Brenner, *Nat. Commun.* **6**, 6203 (2015)
  - [55] T.E. Ouldridge, A.A. Louis, J.P.K. Doye, *J. Phys.: Condens. Matter* **22**, 104102 (2010)
  - [56] T.E. Ouldridge, *J. Chem. Phys.* **137**, 144105 (2012)
  - [57] J. SantaLucia, Jr., D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004)
  - [58] D.H. Turner, D.H. Matthews, *Nucl. Acids Res.* **38**(Database issue), D280 (2010)
  - [59] A.E. Borujeni, H.M. Salis, *J. Am. Chem. Soc.* **138**, 7016 (2016)
  - [60] R.M. Dirks, J.S. Bois, J.M. Schaeffer, E. Winfree, N.A. Pierce, *SIAM Rev.* **29**, 65 (2007)
  - [61] J.V.I. Timonen, M. Latikka, L. Leibler, R.H.A.R.O. Ikkala, *Science* **341**, 253 (2013)
  - [62] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell, 4th ed.* (Garland Science, New York, 2002)
  - [63] D.K.L. J. Paijmans, P.R. ten Wolde, arXiv:1612.02715
  - [64] P. Mehta, D.J. Schwab, *Proc. Nat. Acad. Sci. USA* **109**, 17978 (2012)
  - [65] P. Mehta, A.H. Lang, D.J. Schwab, *J. Stat. Phys.* **162**, 1153 (2016)
  - [66] C.C. Govern, P.R. ten Wolde, *Phys. Rev. Lett.* **113**, 258102 (2014)
  - [67] C.C. Govern, P.R. ten Wolde, *Proc. Nat. Acad. Sci. USA* **111**, 17486 (2014). doi:10.1073/pnas.1411524111
  - [68] J. Stricker, S. Cookson, M. Bennett, W. Mather, L. Tsimring, J. Hasty, *Nature* **454**, 516 (2008)
  - [69] D.Y. Zhang, G. Seelig, *Nat. Chem.* **3**, 103 (2011)
  - [70] N.G. Van Kampen, *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam, 2007)
  - [71] U. Seifert, *Eur. Phys. J. E* **34**, 26 (2011)
  - [72] T.L. Hill, *Free energy transduction and biochemical cycle kinetics, 2nd ed.* (Dover, Mineola, NY, 1989)
  - [73] D. Zhang, E. Winfree, *J. Am. Chem. Soc.* **131**, 17303 (2009)
  - [74] Y.J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Solveichik, G. Seelig, *Nat. Nanotechnol.* **8**, 755 (2013)
  - [75] N. Srinivas, T.E. Ouldridge, P. Šulc, J. Schaeffer, B. Yurke, A.A. Louis, J.P.K. Doye, E. Winfree, *Nucl. Acids Res.* **41**, 10641 (2013)
  - [76] S. Reuveni, I. Meilijson, M. Kepiec, E. Rupp, T. Tuller, *PLoS Comput. Biol.* **7**, e1002127 (2011)
  - [77] C. Flamm, W. Fontana, I.L. Hofacker, P. Schuster, *RNA* **6**, 325 (2000)
  - [78] J.M. Schaeffer, C. Thachuk, E. Winfree, *Lect. Notes Comput. Sc.* **9211**, 194 (2015)
  - [79] F. Dannenberg, K.E. Dunn, J. Bath, M. Kwiatkowska, A.J. Turberfield, T.E. Ouldridge, *J. Chem. Phys.* **143**, 165102 (2015)
  - [80] A.J. Turberfield, J.C. Mitchell, B. Yurke, A.P. Mills, M.I. Blakey, F.C. Simmel, *Phys. Rev. Lett.* **90**(11), 118102 (2003)

- [81] P. Yin, H.M. Choi, C.R. Calvert, N.A. Pierce, *Nature* **451**, 318 (2008)
- [82] H.M.T. Choi, V.A. Beck, N.A. Pierce, *ACS Nano* **8**, 4284 (2014)
- [83] W. Meng, R. Muscat, M.L. McKee, A.H.E. P. J. Milnes, J. Bath, B.G. Davis, T. Brown, R.K. O'Reilly, A.J. Turberfield, *Nat. Chem.* **8**, 542 (2016)
- [84] Y. Gao, L.K. Wolf, R.M. Georgiadis, *Nucl. Acids Res.* **34**, 3370 (2006)
- [85] J.S. Schreck, T.E. Ouldrige, . Romano, P. Šulc, L.P. Shaw, A.A. Louis, J.P. Doye, *Nucl. Acids Res.* **43**, 6181 (2015)
- [86] R.R.F. Machinek, T.E. Ouldrige, N.E.C. Haley, J. Bath, A.J. Turberfield, *Nat. Commun.* **5**, 5324 (2014)
- [87] M. Lakin, S. Youssef, L.Cardelli, A. Phillips, *J. R. Soc. Interface* **9**, 470 (2011)
- [88] K.A. Dill, J.L. MacCallum, *Science* **338**, 1042 (2012)
- [89] K.E. Dunn, F. Dannenberg, T.E. Ouldrige, M. Kwiatkowska, A.J. Turberfield, J. Bath, *Nature* **525**, 82 (2015)
- [90] A. Reinhardt, D. Frenkel, *Phys. Rev. Lett.* **112**, 238103 (2014)
- [91] J. Song, J.M. Arbona, Z. Zhang, L. Liu, E. Xie, J. Elezgaray, J.P. Aime, K.V. Gothelf, F. Besenbacher, M. Dong, *J. Am. Chem. Soc.* **134**, 9844 (2012)
- [92] C.H. Bennett, *Int. J. Theor. Phys.* **21**, 905 (1982)
- [93] R. Cummings, D. Doty, D. Soloveichik, *Lect. Notes Comput. Sc.* **9211**, 37 (2015)
- [94] H.L. Chen, D. Doty, D. Soloveichik, *Nat. Comput.* **7433**, 25 (2012)
- [95] C. Briat, A. Gupta, M. Khammash, *Cell Systems* **2**, 15 (2016)
- [96] L. Qian, D. Soloviechick, E. Winfree, *DNA 16, LNCS* **6518**, 123 (2011)
- [97] C.J. Adkins, *An introduction to thermal physics* (Cambridge University Press, Cambridge, UK, 1987)
- [98] R. Landauer, *IBM J. Res. Dev.* **5**, 183 (1961)
- [99] T. Sagawa, M. Ueda, *Phys. Rev. Lett.* **102**, 250602 (2009)
- [100] C. Thachuck, A. Condon, *Lect. Notes Comput. Sc.* **7433**, 135 (2012)
- [101] C. Thachuk, *Lect. Notes Comput. Sc.* **7948**, 247 (2013)
- [102] T.E. Ouldrige, C.C. Govern, P.R. ten Wolde, *arXiv:1503.00909*
- [103] T.E. Ouldrige, P.R. ten Wolde, *arXiv:1609.05554*
- [104] H. Qian, *Annu. Rev. Phys. Chem.* **58**, 113 (2007)
- [105] D.A. Beard, H. Qian, *Chemical Biophysics* (Cambridge University Press, 2008)
- [106] J.J. Hopfield, *Proc. Natl. Acad. Sci. USA* **71**, 4135 (1974)
- [107] J. Ninio, *Biochimie* **57**(5), 587 (1975)
- [108] F. Tostevin, M. Howard, *Phys. Biol.* **3**, 1 (2006)
- [109] J. Bath, S.J. Green, A.J. Turberfield, *Angew. Chem. Int. Ed.* **117**, 4432 (2005)
- [110] J. Bath, S.J. Green, K.E. Allan, A.J. Turberfield, *Small* **5**, 1513 (2009)
- [111] S.F.J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, A.J. Turberfield, *Nat. Nanotechnol.* **6**, 166 (2011)
- [112] Y. Tian, Y. He, Y. Chen, P. Yin, C. Mao, *Angew. Chem. Int. Ed.* **44**, 4355 (2005)
- [113] K. Lund, A.J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M.N. Stojanovic, N.G. Walter, E. Winfree, H. Yan, *Nature* **465**, 206 (2010)
- [114] T.G. Cha, J. Pan, H. Chen, J. Salgado, X. Li, C. Mao, J.H. Choi, *Nat. Nanotechnol.* **9**, 39 (2014)
- [115] R.A. Muscat, J. Bath, A.J. Turberfield, *Nano Lett.* **11**, 982 (2011)
- [116] S.J. Green, J. Bath, A.J. Turberfield, *Phys. Rev. Lett.* **101**, 238101 (2008)
- [117] T. Omabegho, R. Sha, N.C. Seeman, *Science* **324**, 67 (2009)
- [118] J.M. Parrondo, J.M. Horowitz, T. Sagawa, *Nat. Phys.* **11**, 131 (2015)
- [119] M. Esposito, C. Van den Broeck, *EPL-Europhysics Letters* **95**, 40004 (2011)
- [120] P. Sartori, S. Pigolotti, *Phys. Rev. X* **5**, 041039 (2015)
- [121] M. Nguyen, S. Vaikuntanathan, *Proc. Nat. Acad. Sci. USA* **113**, 14213 (2016)
- [122] D. Collin, F. Ritort, C. Jarzynski, S.B. Smith, I. Tinoco, Jr, C. Bustamante, *Nature* **437**, 231 (2005)
- [123] M. Engel, D.B. Ritchie, D.A.N. Foster, K.S.D. Beach, M.T. Woodside, *Phys. Rev. Lett.* **113**, 238104 (2014)
- [124] G. Lan, P. Sartori, S. Neumann, V. Sourjik, Y. Tu, *Nature Physics* **8**, 422 (2012)
- [125] A.C. Barato, U. Seifert, *Phys. Rev. Lett.* **114**, 158101 (2015)
- [126] P. Pietzonka, A.C. Barato, U. Seifert, *arXiv:1609.08046*
- [127] J.L. England, *The Journal of Chemical Physics* **139**(12), 121923 (2013)
- [128] A.C. Barato, D. Hartich, U. Seifert, *New. J. Phys.* **16**, 103024 (2014)
- [129] S. Ito, T. Sagawa, *Nat. Comm.* **6**, 7498 (2015)
- [130] O. Maroney, *Stud. Hist. Philos. M. P.* **36**, 355 (2005)
- [131] J. Ladyman, S. Presnell, A.J. Short, B. Groisman, *Stud. Hist. Philos. M. P.* **38**, 58 (2007)
- [132] J.D. Norton, *Stud. Hist. Philos. M. P.* **42**, 184 (2011)
- [133] J. Dunkel, S. Hilbert, *Nat. Phys.* **10**, 67 (2014)
- [134] D. Mandal, C. Jarzynski, *Proc. Nat. Acad. Sci. USA* **109**, 11641 (2012)
- [135] D. Mandal, H.T. Quan, C. Jarzynski, *Phys. Rev. Lett.* **111**, 030602 (2013)
- [136] T. McGrath, N.S. Jones, T.E. ten Wolde, P. R. and Ouldrige, *Phys. Rev. Lett.* **118**, 028101 (2017)
- [137] M. Gopalkrishnan, *Personal Communication*
- [138] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz, *Nature* **483**(7388), 187 (2012)
- [139] Y. Jun, M. Gavrilov, J. Bechhoefer, *Phys. Rev. Lett.* **113**, 190601 (2014)